Análisis y diseño de experimentos

HUMBERTO GUTIÉRREZ PULIDO - ROMÁN DE LA VARA SALAZAR



Mc Graw Hill

SEGUNDA EDICIÓN

Análisis y diseño de experimentos

Análisis y diseño de experimentos

Segunda edición

Humberto Gutiérrez Pulido
Centro Universitario de Ciencias Exactas e Ingeniería
Universidad de Guadalajara

Román de la Vara Salazar
Centro de Investigación de Matemáticas
Guanajuato, México

Revisión técnica:

Adolfo Cano Carrasco
Departamento de Ingeniería Industrial
Instituto Tecnológico de Sonora

Mucio Osorio Sánchez
Departamento de Matemáticas
Instituto Tecnológico de Sonora

MÉXICO • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA
LISBOA • MADRID • NUEVA YORK • SAN JUAN • SANTIAGO
AUCKLAND • LONDRES • MILÁN • MONTREAL • NUEVA DELHI
SAN FRANCISCO • SINGAPUR • SAN LUIS • SIDNEY • TORONTO

Director Higher Education: Miguel Ángel Toledo Castellanos
Director editorial: Ricardo A. del Bosque Alayón
Editor sponsor: Pablo E. Roig Vázquez
Editora de desarrollo: Lorena Campa Rojas
Supervisor de producción: Zeferino García García

Diseño de portada: Jorge Matías-Garnica / Brenda Rodríguez

ANÁLISIS Y DISEÑO DE EXPERIMENTOS
Segunda edición

Prohibida la reproducción total o parcial de esta obra,
por cualquier medio, sin la autorización escrita del editor.

McGraw-Hill Interamericana

DERECHOS RESERVADOS © 2008, respecto a la segunda edición en español por
McGRAW-HILL/INTERAMERICANA EDITORES, S.A. de C.V.
A Subsidiary of ***The McGraw-Hill*** *Companies, Inc*.
Edificio Punta Santa Fe
Prolongación Paseo de la Reforma 1015, Torre A
Piso 17, Colonia Desarrollo Santa Fe,
Delegación Álvaro Obregón
C.P. 01376, México, D.F.
Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

ISBN-10: 970-10-6526-3
ISBN-13: 978-970-10-6526-6
(ISBN: 970-10-4017-1 edición anterior)

1234567890 09865432107

Impreso en México *Printed in Mexico*

Dedicatoria

A Irma, Arnoldo y Noel
H.G.P.

A Rosalinda y Armida
R.V.S.

Contenido

Acerca de los autores XI
Agradecimientos XIII
Prefacio XV

CAPÍTULO 1 Introducción al diseño de experimentos 2
El diseño de experimentos hoy 4
Definiciones básicas en el diseño de experimentos 6
Etapas en el diseño de experimentos 10
Consideraciones prácticas sobre el uso de métodos estadísticos 12
Principios básicos 12
Clasificación y selección de los diseños experimentales 14

CAPÍTULO 2 Elementos de inferencia estadística: experimentos con uno y dos tratamientos 18
Población y muestra, parámetros y estadísticos 20
Distribuciones de probabilidad e inferencia 21
Estimación puntual y por intervalo 23
Conceptos básicos de prueba de hipótesis 29
Planteamiento de una hipótesis estadística 30
Prueba para la media 34
Prueba para la varianza 36
Tres criterios de rechazo o aceptación equivalentes 37
Hipótesis para dos medias: comparación de dos tratamientos 39
Prueba para la igualdad de varianzas 43
Poblaciones pareadas (comparación de dos medias con muestras dependientes) 44
Resumen de fórmulas para procedimientos de prueba de hipótesis 49
Uso de un software estadístico 49

CAPÍTULO 3 Experimentos con un solo factor (análisis de varianza) 60
Diseño completamente al azar y ANOVA 62
Comparaciones o pruebas de rango múltiples 74
Verificación de los supuestos del modelo 81
Elección del tamaño de la muestra 89
Uso de software computacional 91

CAPÍTULO 4 Diseños de bloques 100
Diseño de bloques completos al azar 102
Diseño en cuadro latino 109

Diseño en cuadro grecolatino 115
Uso de software 116

CAPÍTULO 5 Diseños factoriales 126
Conceptos básicos en diseños factoriales 128
Experimentación factorial *vs.* mover un factor a la vez 132
Diseños factoriales con dos factores 134
Diseños factoriales con tres factores 143
Transformaciones para estabilizar varianza 149
Diseño factorial general 150
Modelos de efectos aleatorios 153
Cómo hacerlo con software 157

CAPÍTULO 6 Diseños factoriales 2^k 166
Diseño factorial 2^2 168
Experimento 2^2: ejemplo integrador 173
Diseño factorial 2^3 183
Experimento 2^3: ejemplo integrador 185
Diseño factorial general 2^k 192
Diseño factorial 2^k no replicado 195
Experimento 2^5 no replicado: ejemplo integrador 199
Cuando la significancia de los efectos es menos clara: un ejemplo 208
Factoriales 2^k con punto al centro 214
Factoriales 2^k en bloques 216
Uso de software estadístico 220

CAPÍTULO 7 Diseños factoriales 3^k y factoriales mixtos 236
Diseños factoriales 3^k 238
Factoriales mixtos 247
Uso de software estadístico 250

CAPÍTULO 8 Diseños factoriales fraccionados 2^{k-p} 258
Diseño factorial fraccionado 2^{k-1} 261
El concepto de resolución 265
Construcción de fracciones 2^{k-1} 265
Experimento 2^{5-1}: ejemplo integrador 266
Diseños factoriales fraccionados 2^{k-2} 269
Diseño factorial fraccionado 2^{k-p} 273
Experimento 2^{7-4}: ejemplo integrador 276
Tópicos adicionales sobre factoriales fraccionados 279
Uso de software 285

CAPÍTULO 9 Introducción al diseño robusto (Taguchi) 294
Filosofía Taguchi 296
El concepto de robustez 298
Factores de control, de ruido y de señal 299
Arreglos ortogonales 303

Diseño con arreglo interno y externo (diseño de parámetros) 307
Razón señal/ruido 307
Uso de software 312

CAPÍTULO 10 Planeación de un experimento 318
Experimentación: una estrategia para probar conjeturas y generar aprendizaje 320
El diseño de experimentos y el ciclo de Deming 322
Etapas y actividades de la planeación y análisis de un experimento . . . 323
Control de factores de bloque y de ruido 330
Qué sigue después del primer experimento 331
Qué hacer cuando ningún efecto es significativo 334

CAPÍTULO 11 Análisis de regresión 338
Regresión lineal simple 340
Pruebas de hipótesis en la regresión lineal simple 346
Calidad del ajuste en regresión lineal simple 351
Estimación y predicción por intervalo en regresión simple 357
Regresión lineal múltiple 360
Pruebas de hipótesis en regresión lineal múltiple 365
Intervalos de confianza y predicción en regresión múltiple 371
Uso de un software estadístico 373

CAPÍTULO 12 Optimización de procesos con metodología de superficie de respuesta 384
Introducción a la metodología de superficie de respuesta 386
Técnicas de optimización 393
Diseños de superficie de respuesta 413
Uso de software estadístico 420

CAPÍTULO 13 Optimización simultánea de varias respuestas 432
Optimización simultánea 434
Método gráfico 436
Método de la función de deseabilidad 441
Trabajo con un software estadístico 446

CAPÍTULO 14 Diseños anidados y diseños en parcelas divididas . . . 452
Diseños anidados 454
Modelo y análisis estadístico del diseño anidado 455
Diseños en parcelas divididas 460
Modelo y análisis estadístico de los diseños en parcelas divididas 465
Cómo hacer los cálculos usando software 473

CAPÍTULO 15 Diseño de experimentos con mezclas 480
El problema del diseño de experimentos con mezclas 482
Algunos diseños de mezclas y sus modelos estadísticos 486

Ajuste del modelo y caracterización de la superficie de respuesta 490
Restricciones en los componentes de una mezcla 494
Uso de software estadístico 500

APÉNDICE A Tablas 509

APÉNDICE B Uso de sistemas computacionales 521
Etapas al planear y analizar un experimento en un paquete estadístico 522
Sistema *Statgraphics* 522
Sistema *Minitab* 526
Sistema *JMP*. 528
Diseño de experimentos usando *SPSS* 529

REFERENCIAS Y BIBLIOGRAFÍA. 533

ÍNDICE ANALÍTICO. 537

Acerca de los autores

Humberto Gutiérrez Pulido es miembro del Sistema Nacional de Investigadores y profesor investigador en el Centro Universitario de Ciencias Exactas e Ingenierías de la Universidad de Guadalajara. Obtuvo el doctorado en estadística por el Centro de Investigación en Matemáticas (CIMAT) de Guanajuato, México.

Entre las empresas en las que ha impartido capacitación y/o asesoría en calidad total, control estadístico y diseño de experimentos, destacan las siguientes: Cervecería Modelo, Tequila Herradura, Kodak, Hitachi, Jabil, Coca-Cola, Sanmina-SCI y Colcafé.

A lo largo de su trayectoria profesional ha escrito siete libros, cinco de ellos publicados por McGraw-Hill, y 50 artículos de investigación. Asimismo, ha sido conferenciante a nivel nacional e internacional.

Román de la Vara Salazar es investigador en el Centro de Investigación en Matemáticas (CIMAT) de Guanajuato, México.

Es doctor en estadística por el CIMAT.

Entre las empresas en las que ha impartido capacitación y/o asesoría en ingeniería para la calidad y estadística, destacan las siguientes: Pemex, INEGI, Motorola, Comisión Federal de Electricidad, CENAM, Mabe, General Motors y Kodak.

Agradecimientos

Agradecer a todas las personas que directa o indirectamente contribuyeron en un libro como éste no es fácil, ya que a lo largo de los años se acumulan ideas, comentarios, dudas, ejemplos, datos, respuestas, discusiones y experiencia de las personas con las que se ha tenido contacto en el terreno profesional. Al final, las respuestas a esas dudas, las experiencias y las diferentes aportaciones se han vertido en los 15 capítulos del libro, alrededor de 50 ejemplos y más de 300 preguntas y ejercicios reales. De cualquier manera, y disculpándonos de antemano por las omisiones, quisiéramos dejar testimonio de nuestro agradecimiento a las siguientes personas: Julio Yuen (Flextronics); Leopoldo Torres y Gustavo Pacheco (Hitachi); Miguel Cedeño (Tequila Herradura); Ignacio Gallo, Jorge Quirarte (Kodak); Martín Marín (Sanmina SCI); Oscar Famoso y Pedro Ponce (Panasonic); Alberto Godínez, Felipe Camacho y Daniel Romo (Lucent Technology); Armando Miramontes (Innopack); Gustavo Basurto (Coca-Cola); Enrique Villa, Jorge Domínguez y Gustavo Torres (CIMAT); Porfirio Gutiérrez, Osvaldo Camacho, Lizbeth Díaz, Agustín Rodríguez, María de Jesús Guzmán, Cecilia Garibay (Universidad de Guadalajara); Víctor Aguirre (ITAM); Javier Quezada (Tec. de Monterrey); Jorge Villa (Universidad de Sonora); Cuauhtémoc Reyes (Universidad Autónoma de Sinaloa); Edmundo Dávila y Rubén Cárdenas (Jabil Circuits); Joaquín Ávalos (Cervecería Modelo); Joel Cárdenas (Pemex); José Toro (Mabe Sanyo); Valentín Gutiérrez (Emerson), y Ángela Atilano (Colcafé).

Deseamos agradecer la valiosa participación de los siguientes profesores en la presente edición:

José Humberto Loría Arcila, *Universidad Autónoma de Yucatán*
Sonia Avilés Ortiz, *Instituto Tecnológico de Sonora*
Elízabeth González Valenzuela, *Instituto Tecnológico de Sonora*
Rocío Juárez, *Instituto Tecnológico de Sonora*
María Elena Anaya Pérez, *Universidad de Sonora*
Dagoberto Rosas Pandura, *Instituto Tecnológico de Hermosillo*
Esteban Burguete Hernández, *Universidad de las Américas Puebla*

Prefacio

Este libro es resultado de más de 15 años de enseñanza, capacitación y asesoría sobre diseño y análisis de experimentos. Además de la contribución de los autores, esta obra ha sido posible gracias a las ideas, comentarios, dudas, ejemplos, datos, respuestas, discusiones y experiencia de las personas con las que se ha tenido contacto en el terreno profesional; desde estudiantes universitarios, estudiantes de posgrado, investigadores, así como personal técnico y directivo de empresas. Las respuestas a esas dudas, las experiencias y los diferentes aportes se virtieron en los 15 capítulos de este libro, alrededor de 50 ejemplos y más de 300 preguntas y ejercicios reales.

Esperamos que esta obra resulte una contribución para enfrentar de mejor manera los tiempos actuales, ya que la globalización y la alta competencia son una realidad tan contundente que deja poco lugar a dudas acerca de la necesidad de que las organizaciones enfrenten de manera rápida y eficaz esta competencia. Es posible afirmar que la globalización ha dejado atrás muchas discusiones en nuestros países sobre la forma de enfrentar los nuevos tiempos. En la actualidad se sabe, en casi todas las empresas y organizaciones, que ya no hay clientes cautivos y que en cualquier momento los clientes pueden encontrar una mejor alternativa. En este contexto, en la investigación y en los procesos de producción es necesario mejorar la calidad de los productos, reducir costos, reducir tiempos de ciclo, diseñar, rediseñar o hacer cambios en los procesos, sustituir materiales, modificar métodos, diseñar productos y procesos robustos, etc. Una de las metodologías clave para que todo esto se haga eficazmente es el diseño y análisis estadístico de experimentos.

El papel crucial que en la actualidad juega el diseño de experimentos en los centros de investigación y en los procesos productivos se fue consolidando a lo largo de la segunda mitad del siglo XX, y particularmente en los años ochenta recibió un impulso decisivo debido a la influencia del control de calidad en Japón, en donde se dieron cuenta de que más que detectar la mala calidad es mejor enfocar esfuerzos de prevención. En los últimos 20 años el diseño de experimentos se fue consolidando hasta convertirse, hoy en día, en una herramienta fundamental en las tareas del personal técnico de prácticamente todo tipo y tamaño de industria. Esto lo ha reconocido la mayoría de los centros educativos que tienen como tarea la formación de ingenieros, biólogos, químicos, agrónomos, etc.; asimismo, han incorporado el diseño de experimentos como parte de su formación básica. En muchas industrias este tipo de profesionistas reciben un entrenamiento adicional en diseño de experimentos, a fin de fortalecer sus conocimientos y habilidades en este campo para que estén en posibilidades de generar más y mejores acciones de perfeccionamiento e innovación.

En este libro se describen los aspectos más importantes del diseño y análisis de experimentos, y aunque se ven los aspectos matemáticos, se hace énfasis en los conceptos, así como en cuándo aplicar cada tipo de diseño, cómo aplicarlo y cómo hacer el análisis e interpretación de los datos obtenidos mediante el experimento. Nos hemos apoyado en muchos ejemplos para resaltar los aspectos finos de la aplicación del diseño y análisis de experimentos; de esta forma hacemos evidente la gran utilidad que tienen como herramienta de mejora e innovación.

A lo largo de más de 15 años en que hemos enseñado y asesorado la aplicación del diseño de experimentos, hemos visto que la gente con formación técnica capta rápidamente su utilidad y, cuando los aplica, pronto se familiariza y los incorpora a su "caja" de herramientas de uso relativamente cotidiano. La gente que está ligada a la investigación, la operación y la mejora de los procesos está acostumbrada a realizar pruebas (experimentos) de diferente índole, por lo que en este caso el reto es que estas pruebas se hagan en forma planeada y óptima (diseño estadístico de las pruebas), en lugar de la tradicional "prueba y error". En cada capítulo se describe la forma en que se puede utilizar un software estadístico para facilitar tanto el diseño como el análisis de cada tipo de esquema experimental. Al final de cada capítulo se plantean preguntas y ejercicios como material complementario para ayudar al profesor y al alumno, y en general al usuario, a lograr un mejor aprendizaje de la técnica.

El libro cubre prácticamente todo el material de cualquier curso de diseño de experimentos a nivel licenciatura y posgrado. También incluye tópicos adicionales que en ocasiones no se alcanzan a cubrir en un solo curso, pero que es necesario conocer para lograr una mejor panorámica de las posibilidades de aplicación del diseño de experimentos. Tal es el caso de los contenidos de los últimos seis capítulos.

Sobre la segunda edición

En esta nueva edición se realizó una revisión detallada con el propósito de hacer más fácil y clara la lectura de la obra, ampliar algunos temas, agregar ejercicios, eliminar redundancias y mejorar la definición de conceptos clave. Los mayores cambios resultaron en los capítulos: 1, Introducción al diseño; 2, Elementos de inferencia estadística; 3, Análisis de varianza; 4, Diseños en bloques, y el 12, Superficie de respuesta. Además, para que el libro cubra mejor las situaciones experimentales que se presentan en un contexto de investigación y en los problemas industriales, se agregaron tres nuevos capítulos que incluyen tópicos del diseño y análisis de experimentos que cada día tienen mayor aplicación, tales como la optimización multirrespuesta, los diseños anidados y en parcelas divididas y el diseño de experimentos con mezclas.

Al final de cada capítulo se incluyeron breves explicaciones de cómo utilizar algunos sistemas computacionales para hacer el diseño y análisis de los experimentos. En este sentido se hizo especial énfasis en los programas *Statgraphics* y *Minitab*. Además, en algunos capítulos se agregó información acerca de *Excel* y *Design Expert*.

Esperamos que con estas mejoras y nuevos materiales el libro siga siendo bien recibido por la comunidad Iberoamericana.

Análisis y diseño de experimentos

Capítulo 1
Introducción al diseño de experimentos

Sumario

- El diseño de experimentos hoy
- Definiciones básicas en el diseño de experimentos
- Etapas en el diseño de experimentos
- Consideraciones prácticas sobre el uso de métodos estadísticos
- Principios básicos
- Clasificación y selección de los diseños experimentales

Objetivos de aprendizaje

- Conocer el papel fundamental que juega el diseño de experimentos en el mejoramiento de procesos y en la investigación.
- Identificar los principios básicos y la terminología adecuada en el diseño de experimentos.
- Describir las etapas más importantes en la investigación experimental.

Mapa conceptual



Conceptos clave

- Aleatorización
- Bloqueo
- Diseño de experimentos
- Error aleatorio
- Error experimental
- Experimento
- Factores controlables
- Factores estudiados
- Factores no controlables
- Matriz de diseño
- Niveles
- Planeación
- Proceso de deducción
- Proceso de inducción
- Tratamiento
- Repetición
- Unidad experimental
- Variable de respuesta

El diseño de experimentos hoy

En el campo de la industria es frecuente hacer experimentos o pruebas con la intención de resolver un problema o comprobar una idea (conjetura, hipótesis); por ejemplo, hacer algunos cambios en los materiales, métodos o condiciones de operación de un proceso, probar varias temperaturas en una máquina hasta encontrar la que da el mejor resultado o crear un nuevo material con la intención de lograr mejoras o eliminar algún problema.

Sin embargo, es común que estas pruebas o experimentos se hagan sobre la marcha, con base en el ensayo y error, apelando a la experiencia y a la intuición, en lugar de seguir un plan experimental adecuado que garantice una buena respuesta a las interrogantes planteadas. Algo similar ocurre con el análisis de los datos experimentales, donde más que hacer un análisis riguroso de toda la información obtenida y tomar en cuenta la variación, se realiza un análisis informal, "intuitivo". Es tal el poder de la experimentación que, en ocasiones, se logran mejoras a pesar de que el experimento se hizo con base en el ensayo y error. Sin embargo, en situaciones de cierta complejidad no es suficiente aplicar este tipo de experimentación, por lo que es mejor proceder siempre en una forma eficaz que garantice la obtención de las respuestas a las interrogantes planteadas en un lapso corto de tiempo y utilizando pocos recursos.

El *diseño estadístico de experimentos* es precisamente la forma más eficaz de hacer pruebas. El diseño de experimentos consiste en determinar cuáles pruebas se deben realizar y de qué manera, para obtener datos que, al ser analizados estadísticamente, proporcionen evidencias objetivas que permitan responder las interrogantes planteadas, y de esa manera clarificar los aspectos inciertos de un proceso, resolver un problema o lograr mejoras. Algunos problemas típicos que pueden resolverse con el diseño y el análisis de experimentos son los siguientes:

1. Comparar a dos o más materiales con el fin de elegir al que mejor cumple los requerimientos.
2. Comparar varios instrumentos de medición para verificar si trabajan con la misma precisión y exactitud.
3. Determinar los factores (las x vitales) de un proceso que tienen impacto sobre una o más características del producto final.
4. Encontrar las condiciones de operación (temperatura, velocidad, humedad, por ejemplo) donde se reduzcan los defectos o se logre un mejor desempeño del proceso.
5. Reducir el tiempo de ciclo del proceso.
6. Hacer el proceso insensible o robusto a oscilaciones de variables ambientales.
7. Apoyar el diseño o rediseño de nuevos productos o procesos.
8. Ayudar a conocer y caracterizar nuevos materiales.

En general, cuando se quiere mejorar un proceso existen dos maneras básicas de obtener la información necesaria para ello: una es observar o monitorear vía herramientas estadísticas, hasta obtener señales útiles que permitan mejorarlo; se dice que ésta es una estrategia pasiva. La otra manera consiste en experimentar, es decir, hacer cambios estratégicos y deliberados al proceso para provocar dichas señales útiles. Al analizar los resultados del experimento se obtienen las pautas a seguir, que

muchas veces se concretan en mejoras sustanciales del proceso. En este sentido, experimentar es mejor que sentarse a esperar a que el proceso nos indique por sí solo cómo mejorarlo. El *diseño de experimentos* (DDE) es un conjunto de técnicas activas, en el sentido de que no esperan que el proceso mande las señales útiles, sino que éste se "manipula" para que proporcione la información que se requiere para su mejoría.

El saber diseño de experimentos y otras técnicas estadísticas, en combinación con conocimientos del proceso, sitúan al responsable del mismo como un observador perceptivo y proactivo que es capaz de proponer mejoras y de observar algo interesante (oportunidades de mejora) en el proceso y en los datos donde otra persona no ve nada.

Diseño de experimentos
Consiste en planear y realizar un conjunto de pruebas con el objetivo de generar datos que, al ser analizados estadísticamente, proporcionen evidencias objetivas que permitan responder las interrogantes planteadas por el experimentador sobre determinada situación.

Diseño de experimentos en la investigación

Lo que se ha dicho hasta el momento también es válido en el campo de la investigación científica o aplicada, ya que a fin de cuentas, el objetivo es generar nuevas ideas y mejores respuestas a las interrogantes del investigador sobre el objeto de estudio.

El *objetivo de los métodos estadísticos* es lograr que el proceso de generar conocimiento y aprendizaje sea lo más eficiente posible. En este proceso, que ha demostrado ser secuencial, interactúan dos polos (véase figura 1.1), por un lado están la teoría, los modelos, las hipótesis, las conjeturas y los supuestos; por el otro, están la realidad, los hechos, los fenómenos, la evidencia y los datos. Así, como se comenta en Box *et al*. (1978), una hipótesis inicial lleva a un *proceso de deducción* en el que las consecuencias derivadas de la hipótesis pueden ser comparadas con los datos. Cuando las consecuencias y los datos no corresponden, entonces la discrepancia puede llevar a un *proceso de inducción*, en el cual se modifica la hipótesis original. De esta manera inicia un segundo ciclo de la interacción de teoría y datos, en el cual las consecuencias de la hipótesis modificada son comparadas con los datos (los viejos y los que se obtengan en este nuevo ciclo); esto puede llevar a futuras modificaciones y a la obtención de conocimiento.

Proceso de deducción
Es cuando las consecuencias derivadas de la hipótesis pueden ser comparadas con los datos.

Proceso de inducción
Es cuando las consecuencias de la hipótesis original y los datos no están de acuerdo, por lo que se inicia este proceso para cambiar tal hipótesis.

Este proceso interactivo de aprendizaje puede visualizarse como un ciclo de retroalimentación (figura 1.2), en el cual las discrepancias entre los datos y las consecuencias de la hipótesis H_1, llevan a una hipótesis modificada H_2, y de la verificación de ésta, además de conocimiento, se produce una modificación de la modificación (hipótesis H_3) y así sucesivamente.

Breve historia del diseño de experimentos

El diseño estadístico de experimentos, desde su introducción por Ronald A. Fisher en la primera mitad del siglo XX en Inglaterra, se ha utilizado para conseguir un aprendizaje acelerado. El trabajo de Fisher a través de su libro *The Design of Experiments* (1935),

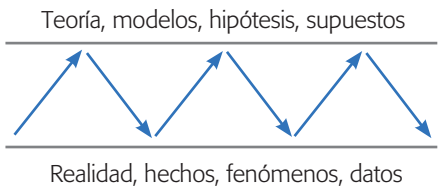


Figura 1.1 Proceso interactivo de la experimentación.

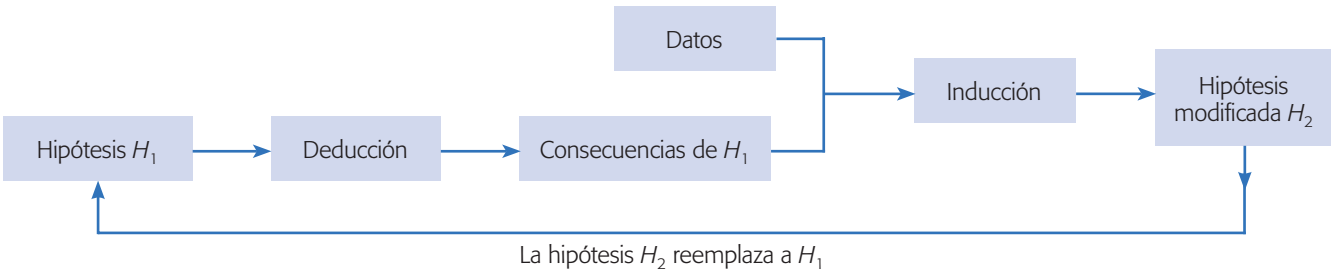


Figura 1.2 El proceso de generación de aprendizaje y conocimiento como un ciclo de retroalimentación.

influyó de manera decisiva en la investigación agrícola, ya que aportó métodos (ahora usados en todo el mundo) para evaluar los resultados de experimentos con muestras pequeñas. La clave de las aportaciones de Fisher radica en que este investigador se dio cuenta de que las fallas en la forma de realizar experimentos obstaculizaba el análisis de los resultados experimentales. Fisher también proporcionó métodos para diseñar experimentos destinados a investigar la influencia simultánea de varios factores.

Los desarrollos posteriores en diseños de experimentos fueron encabezados por George E. P. Box, quien trabajó como estadístico durante ocho años en la industria química en Inglaterra y desarrolló la metodología de superficie de respuestas (véase Box y Wilson, 1951), la cual incluye nuevas familias de diseños y una estrategia para la experimentación secuencial. Es posible afirmar que entre 1950 y 1980, el diseño de experimentos se convirtió en una herramienta de aplicación frecuente, pero sólo en las áreas de investigación y desarrollo. Hasta la década de 1970, la aplicación a nivel planta o procesos de manufactura no estaba generalizada, debido a la falta de recursos computacionales y a que los ingenieros y especialistas en manufactura carecían de formación en el área de estadística.

En la década de 1980 se dio un gran impulso al conocimiento y la aplicación del diseño de experimentos debido al éxito en calidad de la industria japonesa. El movimiento por la calidad, encabezado por los gurúes Deming e Ishikawa, promovió el uso de la estadística en calidad, donde el diseño de experimentos demostró su utilidad tanto para resolver problemas de fondo como para diseñar mejor los productos y los procesos. En Japón destaca el trabajo de Genichi Taguchi, cuyos conceptos sobre diseño robusto también tuvieron un impacto significativo en la academia en el mundo occidental. Como respuesta al movimiento por la calidad y la mejora de procesos, las industrias empezaron a entrenar a sus ingenieros en la aplicación del diseño de experimentos. Esto continúa en la actualidad; incluso, en los últimos veinte años, las universidades han incorporado el diseño de experimentos como materia obligatoria u operativa en la mayoría de las ingenierías.

Definiciones básicas en el diseño de experimentos

El *diseño de experimentos* es la aplicación del método científico para generar conocimiento acerca de un sistema o proceso, por medio de pruebas planeadas adecuadamente. Esta metodología se ha ido consolidando como un conjunto de técnicas

estadísticas y de ingeniería, que permiten entender mejor situaciones complejas de relación causa-efecto.

Experimento

Un *experimento* es un cambio en las condiciones de operación de un sistema o proceso, que se hace con el objetivo de medir el efecto del cambio sobre una o varias propiedades del producto o resultado. Asimismo, el experimento permite aumentar el conocimiento acerca del sistema. Por ejemplo, en un proceso químico se pueden probar diferentes temperaturas y presiones, y medir el cambio observado en el rendimiento (yield, ppm, defectivo) del proceso. Al analizar los efectos (datos) se obtiene conocimiento acerca del proceso químico, lo cual permite mejorar su desempeño.

Experimento
Es un cambio en las condiciones de operación de un sistema o proceso, que se hace con el objetivo de medir el efecto del cambio en una o varias propiedades del producto o resultado.

Unidad experimental

La *unidad experimental* es la pieza(s) o muestra(s) que se utiliza para generar un valor que sea representativo del resultado del experimento o prueba. En cada diseño de experimentos es importante definir de manera cuidadosa la unidad experimental, ya que ésta puede ser una pieza o muestra de una sustancia o un conjunto de piezas producidas, dependiendo del proceso que se estudia. Por ejemplo, si se quiere investigar alternativas para reducir el porcentaje de piezas defectuosas, en un proceso que produce muchas piezas en un lapso corto de tiempo, es claro que no sería muy confiable que la unidad experimental fuera una sola pieza, en la cual se vea si en una condición experimental estaba defectuosa o no. Aquí, la unidad experimental será cierta cantidad de piezas que se producen en las mismas condiciones experimentales, y al final se analizará cuántas de ellas están defectuosas y cuántas no.

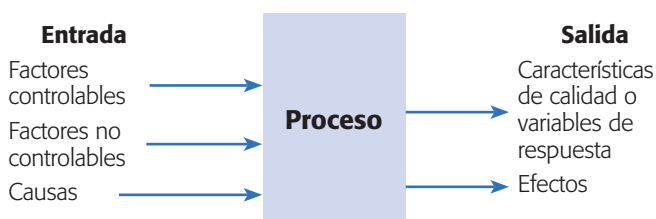
Unidad experimental
Pieza(s) o muestra(s) que se utiliza para generar un valor que sea representativo del resultado de la prueba.

Variables, factores y niveles

En todo proceso intervienen distintos tipos de variables o factores como los que se muestran en la figura 1.3, donde también se aprecian algunas interrogantes al planear un experimento.

Variable(s) de respuesta. A través de esta(s) variable(s) se conoce el efecto o los resultados de cada prueba experimental (véase figura 1.3), por lo que pueden ser características de la calidad de un producto y/o variables que miden el desempeño de

Variable de respuesta
A través de esta(s) variable(s) se conoce el efecto o los resultados de cada prueba experimental.



¿Cuáles características de calidad se van a medir?
¿Cuáles factores controlables deben incluirse en el experimento?
¿Qué niveles debe utilizar cada factor?
¿Cuál diseño experimental es el adecuado?

Figura 1.3 Variables de un proceso y preguntas a responder al diseñar un experimento.

un proceso. El objetivo de muchos estudios experimentales es encontrar la forma de mejorar la(s) variable(s) de respuesta. Por lo general, estas variables se denotan con la letra y.

Factores controlables
Son variables de proceso y/o características de los materiales y los métodos experimentales que se pueden fijar en un nivel dado.

Factores controlables. Son variables de proceso o características de los materiales experimentales que se pueden fijar en un nivel dado. Algunos de éstos son los que usualmente se controlan durante la operación normal del proceso (véase figura 1.3), y se distinguen porque, para cada uno de ellos, existe la manera o el mecanismo para cambiar o manipular su nivel de operación. Esto último es lo que hace posible que se pueda experimentar con ellos. Por ejemplo, si en el proceso se usa agua a 60°C entonces debe existir un mecanismo que permita fijar la temperatura del agua dentro de un rango de operación. Algunos factores o características que generalmente se controlan son: temperatura, tiempo de residencia, cantidad de cierto reactivo, tipo de reactivo, método de operación, velocidad, presión, etc. A los factores controlables también se les llama *variables de entrada*, *condiciones de proceso*, *variables de diseño*, *parámetros del proceso*, las *x de un proceso* o simplemente *factores*.

Factores no controlables
Son variables que no se pueden controlar durante el experimento o la operación normal del proceso.

Factores no controlables o de ruido. Son variables o características de materiales y métodos que no se pueden controlar durante el experimento o la operación normal del proceso. Por ejemplo, algunos factores que suelen ser no controlables son las variables ambientales (luz, humedad, temperatura, partículas, ruido, etc.), el ánimo de los operadores, la calidad del material que se recibe del proveedor (interno o externo). Un factor que ahora es *no controlable* puede convertirse en *controlable* cuando se cuenta con el mecanismo o la tecnología para ello.

Factores estudiados
Son las variables que se investigan en el experimento para observar cómo afectan o influyen en la variable de respuesta.

Factores estudiados. Son las variables que se investigan en el experimento, respecto de cómo influyen o afectan a la(s) variable(s) de respuesta. Los factores estudiados pueden ser controlables o no controlables, a estos últimos quizá fue posible y de interés controlarlos durante el experimento. Para que un factor pueda ser estudiado es necesario que durante el experimento se haya probado en, al menos, dos niveles o condiciones.

En principio, cualquier factor, sea controlable o no, puede tener alguna influencia en la variable de respuesta que se refleja en su media o en su variabilidad. Para fines de un diseño de experimentos deben seleccionarse los factores que se considera, por conocimiento del objeto de estudio, que pueden tener efecto sobre la respuesta de interés. Obviamente, si se decide o interesa estudiar el efecto de un factor no controlable, parte de la problemática a superar durante el diseño es ver la manera en que se controlará durante el experimento tal factor.

Niveles y tratamientos. Los diferentes valores que se asignan a cada factor estudiado en un diseño experimental se llaman *niveles*. Una combinación de niveles de todos los factores estudiados se llama *tratamiento* o *punto de diseño*. Por ejemplo, si en un experimento se estudia la influencia de la velocidad y la temperatura, y se decide probar cada una en dos niveles, entonces cada combinación de niveles (velocidad, temperatura) es un tratamiento. En este caso habría cuatro tratamientos, como se muestra en la tabla 1.1. Es necesario probar cada tratamiento y obtener el correspondiente valor de y.

Tabla 1.1 Puntos de diseño o tratamientos.

Nivel de velocidad	Nivel de temperatura	Tratamiento	*y*
1	1	1	?
2	1	2	
1	2	3	
2	2	4	

De acuerdo con estas definiciones, en el caso de experimentar con un solo factor, cada nivel es un tratamiento.

Error aleatorio y error experimental. Siempre que se realiza un estudio experimental, parte de la variabilidad observada en la respuesta no se podrá explicar por los factores estudiados. Esto es, siempre habrá un remanente de variabilidad que se debe a causas comunes o aleatorias, que generan la variabilidad natural del proceso. Esta variabilidad constituye el llamado *error aleatorio*. Por ejemplo, será parte de este error aleatorio el pequeño efecto que tienen los factores que no se estudiaron, siempre y cuando se mantenga pequeño o despreciable, así como la variabilidad de las mediciones hechas bajo las mismas condiciones. Sin embargo, el error aleatorio también absorberá todos los errores que el experimentador comete durante los experimentos, y si éstos son graves, más que error aleatorio hablaremos de *error experimental*. De predominar éste, la detección de cuáles de los factores estudiados tienen un efecto real sobre la respuesta será difícil, si no es que imposible.

Error aleatorio
Es la variabilidad observada que no se puede explicar por los factores estudiados; resulta del pequeño efecto de los factores no estudiados y del error experimental.

Error experimental
Componente del error aleatorio que refleja los errores del experimentador en la planeación y ejecución del experimento.

Cuando se corre un diseño experimental es importante que la variabilidad observada de la respuesta se deba principalmente a los factores estudiados y en menor medida al error aleatorio, y además que este error sea efectivamente aleatorio. Cuando la mayor parte de la variabilidad observada se debe a factores no estudiados o a un error no aleatorio, no se podrá distinguir cuál es el verdadero efecto que tienen los

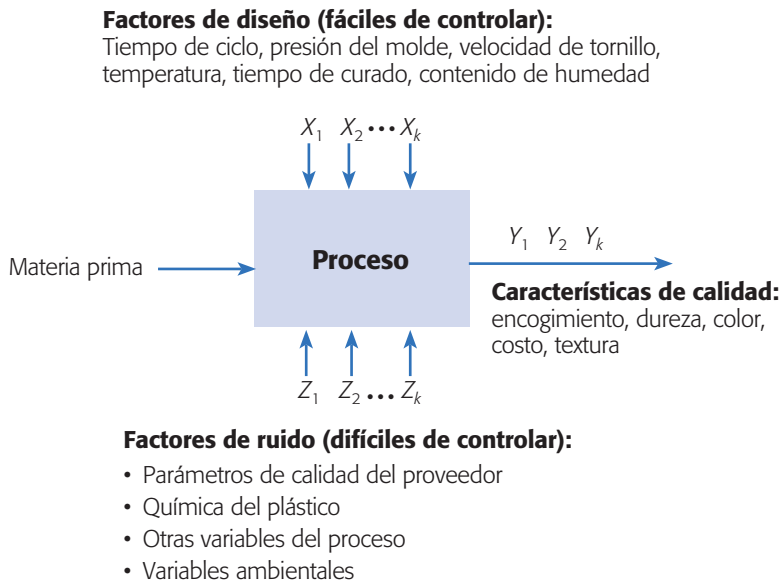

Factores de diseño (fáciles de controlar):
Tiempo de ciclo, presión del molde, velocidad de tornillo, temperatura, tiempo de curado, contenido de humedad
X_1 X_2 ⋯ X_k
Materia prima
Proceso
Y_1 Y_2 Y_k
Características de calidad:
encogimiento, dureza, color, costo, textura
Z_1 Z_2 ⋯ Z_k
Factores de ruido (difíciles de controlar):
• Parámetros de calidad del proveedor
• Química del plástico
• Otras variables del proceso
• Variables ambientales

Figura 1.4 Factores y variables en la fabricación de un envase de plástico.

factores estudiados, con lo que el experimento no alcanzaría su objetivo principal. De aquí la importancia de no dejar variar libremente a ningún factor que pueda influir de manera significativa sobre el comportamiento de la respuesta (principio de bloqueo).

Ejemplo 1.1

En la figura 1.4 se muestran algunas de las variables que intervienen en el proceso de fabricación de un envase de plástico. El problema general es encontrar las condiciones de operación de los factores controlables, que dan por resultado valores óptimos de las características de calidad ahí listadas. También podría ser de interés investigar el efecto de factores no controlables, buscando lograr un proceso insensible (robusto) a su posible efecto. Supongamos que sólo interesa la dureza de la pieza de plástico resultante. Algunas preguntas que se pueden responder con un diseño experimental son: ¿cuáles factores afectan la dureza del plástico?, ¿cómo es que la afectan?, o bien, ¿qué relación hay entre los factores controlables y la dureza?; ¿existen otras condiciones de operación, distintas a las actuales que mejoran la dureza? Estas preguntas se responden probando diferentes combinaciones en los niveles de los factores controlables, seleccionadas de manera adecuada. Esto último significa escoger el diseño experimental más adecuado al problema, que en este caso parece ser un diseño factorial completo o fraccionado (capítulos 5, 6 y 8).

Etapas en el diseño de experimentos

Un aspecto fundamental del diseño de experimentos es decidir cuáles pruebas o tratamientos se van a realizar y cuántas repeticiones de cada uno se requieren, de manera que se obtenga la máxima información al mínimo costo posible. El arreglo formado por los diferentes tratamientos que serán corridos, incluyendo las repeticiones, recibe el nombre de *matriz de diseño* o sólo *diseño.*

Matriz de diseño
Es el arreglo formado por los tratamientos que serán corridos, incluyendo las repeticiones.

Para que un estudio experimental sea exitoso es necesario realizar, por etapas, diferentes actividades. En este sentido, la etapa más importante y a la que se le debe dedicar mayor tiempo es la *planeación* (véase capítulo 10). A continuación se describen de manera breve las etapas del diseño de experimentos con objeto de dar una visión global de lo que implica su correcta aplicación. Varios conceptos que se mencionan en estas etapas se definen con detalle en los siguientes capítulos.

Planeación
Son actividades encaminadas a entender, delimitar el problema u objeto de estudio y seleccionar variables de respuesta y factores. Concluye con la especificación de los tratamientos a realizar y con la organización del trabajo experimental.

Planeación y realización

1. **Entender y delimitar el problema u objeto de estudio.** En la etapa de planeación se deben hacer investigaciones preliminares que conduzcan a entender y delimitar el problema u objeto de estudio, de tal forma que quede claro qué se va a estudiar, por qué es importante y, si es un problema, cuál es la magnitud del mismo.
2. **Elegir la(s) variable(s) de respuesta que será medida en cada punto del diseño y verificar que se mide de manera confiable.** La elección de esta(s) variable(es) es vital, ya que en ella se refleja el resultado de las pruebas. Por ello, se deben elegir aquellas que mejor reflejen el problema o que caractericen al objeto de estudio. Además, se debe tener confianza en

que las mediciones que se obtengan sobre esas variables sean confiables. En otras palabras, se debe garantizar que los instrumentos y/o métodos de medición son capaces de repetir y reproducir una medición, que tienen la precisión (error) y exactitud (calibración) necesaria. Recordemos que los sistemas de medición son la forma en la que percibimos la realidad, por lo que si éstos son deficientes, las decisiones que se tomen con base en ellos pueden ser inadecuadas.

3. **Determinar cuáles factores deben estudiarse o investigarse, de acuerdo a la supuesta influencia que tienen sobre la respuesta.** No se trata de que el experimentador tenga que saber *a priori* cuáles factores influyen, puesto que precisamente para eso es el experimento, pero sí de que utilice toda la información disponible para incluir aquellos que se considera que tienen un mayor efecto.
4. **Seleccionar los niveles de cada factor, así como el diseño experimental adecuado a los factores que se tienen y al objetivo del experimento.** Este paso también implica determinar cuántas repeticiones se harán para cada tratamiento, tomando en cuenta el tiempo, el costo y la precisión deseada.
5. **Planear y organizar el trabajo experimental.** Con base en el diseño seleccionado, organizar y planear con detalle el trabajo experimental, por ejemplo, las personas que van a intervenir, la forma operativa en que se harán las cosas, etc. (véase capítulo 10).
6. **Realizar el experimento.** Seguir al pie de la letra el plan previsto en la etapa anterior, y en caso de algún imprevisto, determinar a qué persona se le reportaría y lo que se haría.

Análisis

En esta etapa no se debe perder de vista que los resultados experimentales son *observaciones muestrales*, no poblacionales. Por ello, se debe recurrir a métodos estadísticos inferenciales para ver si las diferencias o efectos muestrales (experimentales) son lo suficientemente grandes para que garanticen diferencias poblacionales (o a nivel proceso). La técnica estadística central en el análisis de los experimentos es el llamado análisis de varianza ANOVA (acrónimo en inglés).

Interpretación

Aquí, con el respaldo del análisis estadístico formal, se debe analizar con detalle lo que ha pasado en el experimento, desde contrastar las conjeturas iniciales con los resultados del experimento, hasta observar los nuevos aprendizajes que sobre el proceso se lograron, verificar supuestos y elegir el tratamiento ganador, siempre con apoyo de las pruebas estadísticas.

Control y conclusiones finales

Para concluir el estudio experimental se recomienda decidir qué medidas implementar para generalizar el resultado del estudio y para garantizar que las mejoras se mantengan. Además, es preciso organizar una presentación para difundir los logros.

Consideraciones prácticas sobre el uso de métodos estadísticos

En adición a lo dicho en la sección anterior, es importante tomar en cuenta que aunque el uso de metodologías estadísticas por lo general ayuda a hacer más eficiente el proceso de investigación y de solución de problemas, es necesario reconocer que las metodologías estadísticas por sí solas no garantizan investigaciones exitosas, por ello es importante considerar los siguientes puntos:

El conocimiento no estadístico es vital. Para utilizar los métodos estadísticos en general y los diseños de experimentos en particular, en primer lugar se requiere que el experimentador tenga un buen nivel de conocimiento técnico y práctico sobre el fenómeno o proceso que estudia, de tal forma que pueda vislumbrar con cierta facilidad cuáles son los aspectos clave del fenómeno y sea capaz de plantear conjeturas precisas, vislumbrar el tipo de relaciones entre las variables de respuesta y los posibles factores a estudiar. Todo esto ayudará a seleccionar mejor los factores y sus niveles, así como el diseño que es mejor aplicar. Además, ese conocimiento permitirá sacarle un provecho real al análisis estadístico de los resultados y obtener conclusiones que generen aprendizaje y soluciones.

Reconocer la diferencia entre significancia estadística e importancia práctica. En ocasiones, un experimentador puede concluir que dos tratamientos son diferentes estadísticamente, pero que tales diferencias, aunque sean significativas, no necesariamente representan una diferencia que en la práctica sea importante.

Apostarle más a la experimentación secuencial que a un experimento único y definitivo. En ocasiones, los experimentadores novatos pretenden en una sola fase de experimentación contestar todas sus interrogantes sobre un proceso o fenómeno en particular. Sin embargo, esto puede llevar a experimentos muy extensos que consuman demasiados recursos y que retarden la generación de resultados. Por ello es importante considerar como alternativas a diferentes fases de experimentación en forma secuencial, en las cuales se alcance paulatinamente una mayor precisión en los conocimientos y soluciones.

Es importante no confundir la experimentación secuencial con la experimentación a prueba y error (véase sección "Experimentación factorial frente a mover un factor a la vez" del capítulo 5). La *experimentación secuencial* en cada fase sigue una estrategia bien definida y pensada; por lo tanto, en cada fase se obtienen resultados y conclusiones importantes que permiten generar soluciones y conocimiento más refinado para plantear de mejor manera la siguiente fase de experimentación.

Principios básicos

El diseño de experimentos trata de fenómenos que son observables y repetibles. Por lo tanto, sin el pensamiento estadístico, los conceptos de observabilidad y repetibilidad son inherentemente contradictorios. Cualquier cosa observada se aprecia con variabilidad; nada ocurre exactamente de la misma forma dos veces, incluso las me-

diciones del mismo evento varían. Entonces, ¿qué se quiere decir cuando la ciencia demanda que una observación sea repetible?, ¿qué repetición es realmente una repetición?, cuando un resultado es el mismo o difiere, ¿es confirmación o contradicción? Estas preguntas no pueden ser contestadas de manera coherente sin el pensamiento estadístico; por ejemplo, alguien da una nueva receta de chocolate, dice que no falla, pero se prueba y no sale, mientras que el segundo y tercer intento sí funcionan. ¿La receta está comprobada completamente? (Los chocolates tienen más de 800 ingredientes individuales que pueden ser separados.)

De acuerdo con lo anterior, se debe ser muy cuidadoso en la planeación y el análisis de un experimento. El punto de partida para una correcta planeación es aplicar los principios básicos del diseño de experimentos: aleatorización, repetición y bloqueo, los cuales tienen que ver directamente con que los datos obtenidos sean útiles para responder a las preguntas planteadas, es decir, la validez del análisis de los datos se apoya en estos principios.

Aleatorización. Consiste en hacer las corridas experimentales en orden aleatorio (al azar) y con material también seleccionado aleatoriamente. Este principio aumenta la probabilidad de que el supuesto de independencia de los errores se cumpla, lo cual es un requisito para la validez de las pruebas de estadísticas que se realizan. También es una manera de asegurar que las pequeñas diferencias provocadas por materiales, equipo y todos los factores no controlados, se repartan de manera homogénea en todos los tratamientos. Por ejemplo, una evidencia de incumplimiento o violación de este principio se manifiesta cuando el resultado obtenido en una prueba está muy influenciado por la prueba inmediata anterior.

Aleatorización
Consiste en hacer corridas experimentales en orden aleatorio (al azar); este principio aumenta la posibilidad de que el supuesto de independencia de los errores se cumpla.

Repetición. Es correr más de una vez un tratamiento o una combinación de factores. Es preciso no confundir este principio con medir varias veces el mismo resultado experimental. Repetir es volver a realizar un tratamiento, pero no inmediatamente después de haber corrido el mismo tratamiento, sino cuando corresponda de acuerdo con la aleatorización. Las repeticiones permiten distinguir mejor qué parte de la variabilidad total de los datos se debe al error aleatorio y cuál a los factores. Cuando no se hacen repeticiones no hay manera de estimar la variabilidad natural o el error aleatorio, y esto dificulta la construcción de estadísticas realistas en el análisis de los datos.

Repetición
Es correr más de una vez un tratamiento o combinación de factores.

Bloqueo. Consiste en nulificar o tomar en cuenta, en forma adecuada, todos los factores que puedan afectar la respuesta observada. Al bloquear, se supone que el subconjunto de datos que se obtengan dentro de cada bloque (nivel particular del factor bloqueado), debe resultar más homogéneo que el conjunto total de datos. Por ejemplo, si se quieren comparar cuatro máquinas, es importante tomar en cuenta al operador de las máquinas, en especial si se cree que la habilidad y los conocimientos del operador pueden influir en el resultado. Una posible estrategia de bloqueo del factor operador, sería que un mismo operador realizara todas las pruebas del experimento. Otra posible estrategia de bloqueo sería experimentar con cuatro operadores (cuatro bloques), donde cada uno de ellos prueba en orden aleatorio las cuatro máquinas; en este segundo caso, la comparación de las máquinas quizás es más real. Cada operador es un bloque porque se espera que las mediciones del mismo operador sean más parecidas entre sí que las mediciones de varios operadores.

Bloqueo
Es nulificar o tomar en cuenta en forma adecuada todos los factores que pueden afectar la respuesta observada.

Los principios básicos se entenderán mejor en la medida en que se estudien los ejemplos de los capítulos siguientes. En particular, en la sección "Poblaciones pareadas" del capítulo 2, se presentan los experimentos más simples donde la aplicación de estos principios es evidente.

Clasificación y selección de los diseños experimentales

Existen muchos diseños experimentales para estudiar la gran diversidad de problemas o situaciones que ocurren en la práctica. Esta cantidad de diseños hace necesario saber cómo elegir el más adecuado para una situación dada y, por ende, es preciso conocer cómo es que se clasifican los diseños de acuerdo con su objetivo y su alcance.

Los cinco aspectos que más influyen en la selección de un diseño experimental, en el sentido de que cuando cambian por lo general nos llevan a cambiar de diseño, son:

1. El objetivo del experimento.
2. El número de factores a estudiar.
3. El número de niveles que se prueban en cada factor.
4. Los efectos que interesa investigar (relación factores-respuesta).
5. El costo del experimento, tiempo y precisión deseada.

Estos cinco puntos no son independientes entre sí, pero es importante señalarlos de manera separada, ya que al cambiar cualquiera de ellos generalmente cambia el diseño experimental a utilizar (véase capítulo 10). Con base en algunos de estos cinco puntos es posible clasificar los diseños como lo hacemos a continuación.

El *objetivo del experimento* se utiliza como un criterio general de clasificación de los diseños experimentales, mientras que los otros cuatro puntos son útiles para subclasificarlos. En este sentido, de acuerdo con su objetivo y sin pretender ser exhaustivos, los diseños se pueden clasificar como:

1. Diseños para comparar dos o más tratamientos.
2. Diseños para estudiar el efecto de varios factores sobre la(s) respuesta(s).
3. Diseños para determinar el punto óptimo de operación del proceso.
4. Diseños para la optimización de una mezcla.
5. Diseños para hacer el producto o proceso insensible a factores no controlables.

En la figura 1.5 se muestra la clasificación general de los diseños experimentales de acuerdo con su objetivo. Dentro de cada rama se pueden clasificar de acuerdo al número de factores, al tipo de efectos que se pretende estudiar y según las restricciones existentes. En la misma figura se listan los diseños particulares más representativos de cada rama.

Nótese que los diseños factoriales completos y fraccionados ocupan más de un lugar en la figura 1.5; la razón es que estos diseños son eficaces en diversas situacio-

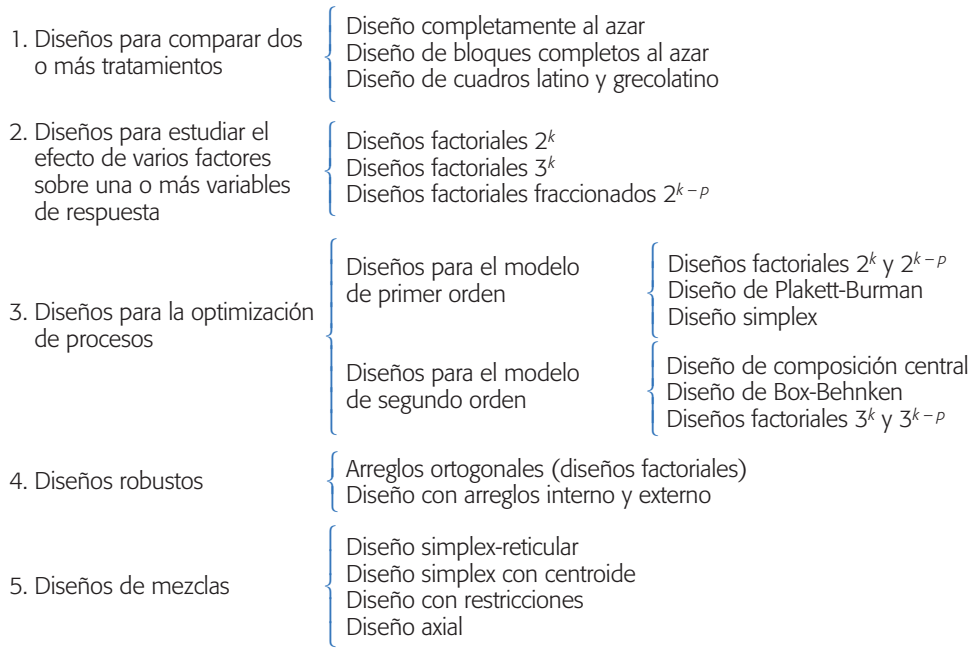

1. Diseños para comparar dos o más tratamientos
Diseño completamente al azar
Diseño de bloques completos al azar
Diseño de cuadros latino y grecolatino
2. Diseños para estudiar el efecto de varios factores sobre una o más variables de respuesta
Diseños factoriales 2^k
Diseños factoriales 3^k
Diseños factoriales fraccionados 2^{k-p}
3. Diseños para la optimización de procesos
Diseños para el modelo de primer orden
Diseños factoriales 2^k y 2^{k-p}
Diseño de Plakett-Burman
Diseño simplex
Diseños para el modelo de segundo orden
Diseño de composición central
Diseño de Box-Behnken
Diseños factoriales 3^k y 3^{k-p}
4. Diseños robustos
Arreglos ortogonales (diseños factoriales)
Diseño con arreglos interno y externo
5. Diseños de mezclas
Diseño simplex-reticular
Diseño simplex con centroide
Diseño con restricciones
Diseño axial

Figura 1.5 Clasificación de los diseños experimentales.

nes prácticas. De hecho, varios de los otros diseños que se mencionan en esta figura son casos particulares o generalizaciones de los diseños factoriales. En los siguientes capítulos se verán con detalle prácticamente todos estos diseños.

Preguntas y ejercicios

1. Explique las ventajas que tiene el diseño de experimentos sobre una estrategia de prueba y error.
2. ¿Qué es un experimento y qué es diseñar un experimento?
3. En el contexto de un diseño de experimentos, ¿qué es una variable de respuesta?, ¿qué es un factor estudiado? y ¿qué relación se esperaría que haya entre la variable y los factores?
4. ¿En un experimento sólo es posible estudiar los factores que actualmente se controlan en la operación normal del proceso?
5. ¿Es posible estudiar cómo influye un factor sobre la variable de respuesta, si el factor se mantiene fijo en todas las corridas o pruebas experimentales? Explique.
6. Se tiene un experimento en el que los factores a estudiar y sus niveles son los siguientes: temperatura (10, 20 y 30°C); tiempo (60 y 90 minutos). Elabore una lista de todos los posibles tratamientos de este diseño.
7. ¿Qué es el error aleatorio y qué es el error experimental?
8. ¿Por qué es importante aleatorizar el orden en que se corren los diferentes tratamientos en un diseño de experimentos?
9. Señale las etapas en el diseño de un experimento, así como algunos aspectos clave de cada una de ellas.

10. ¿Por qué se considera la planeación del experimento como la etapa más importante? Describa cinco actividades que se realizan en esta etapa.

11. Describa de manera breve los tres principios básicos del diseño de experimentos.

12. Explique la diferencia entre significancia práctica y significancia estadística. Proponga un ejemplo donde se tenga la segunda pero no la primera.

13. Describa cinco aspectos que son relevantes al momento de seleccionar el diseño experimental.

14. Mencione dos problemas en su área de trabajo que pudieran abordarse con el diseño de experimentos. Para cada problema enliste algunos factores de control y al menos una variable de respuesta.

15. Suponga que se quiere estudiar el desempeño de un automóvil, y lo que se desea es encontrar los factores que más influyen en su rendimiento. ¿Cuáles podrían ser las variables de respuesta?, ¿cuáles los factores a estudiar?, ¿cuáles los factores no controlables o de ruido?

16. Se quiere comparar el desgaste de dos marcas de llantas A y B, para lo cual se eligen al azar 10 conductores particulares de cierta ciudad. A cinco de ellos, seleccionados al azar, se les instalan gratis las llantas marca A y a los cinco restantes la marca B, con el compromiso por escrito de permitir la verificación del desgaste cada seis meses.
    a) ¿Cree que este experimento permita una comparación justa del desgaste de las dos marcas de llantas?
    b) ¿Qué consideraciones se debieron hacer para lograr una comparación más justa?
    c) Proponga al menos un cambio al experimento que usted considera que mejoraría la comparación.

17. Una compañía farmacéutica realizó un experimento para comprobar los tiempos promedio (en días), que son necesarios para que una persona se recupere de los efectos y las complicaciones que siguen a un resfriado común. En este experimento se compararon a personas que tomaron distintas dosis diarias de vitamina C. Para hacer el experimento se contactó a un número determinado de personas, que en cuanto les daba el resfriado empezaban a recibir algún tipo de dosis. Si la edad de las personas es una posible fuente de variabilidad, explique con detalle cómo aplicaría la idea de bloqueo para controlar tal fuente de variabilidad.

18. En el caso anterior, ¿qué podría pasar si no se controla la posible fuente de variación que es la edad?

19. Un grupo de investigadores trabaja para industrializar la mermelada de tuna; para ello, realizan mermeladas considerando los siguientes factores: *a*) variedad de tuna: tres tipos, *b*) con cáscara o sin cáscara, *c*) completa o la pura pulpa. Por lo tanto, se tienen 12 posibles formas (tratamientos) de producir mermelada.

    La pregunta central que se plantean es si influyen en el sabor los factores considerados, y quisieran encontrar cuál es la mejor combinación de mermelada (tratamiento ganador). Para responder hicieron las 12 combinaciones y pusieron cada una en un recipiente numerado. Enseguida se trasladaban a lugares concurridos donde acomodaban los recipientes ordenados del 1 al 12, y a personas del público les entregaban una hoja de registro y la invitaban a que en el orden dado probaran en pequeñas porciones las mermeladas y anotaran qué tan buena les parecía la mermelada (en una calificación entre 0 a 10). Al final se tuvo la respuesta de 420 personas, donde cada una daba 12 calificaciones (una para cada mermelada). ¿Hay algo que desde su punto de vista invalide los resultados obtenidos? Utilice el sentido común y argumente su respuesta.

Capítulo 2

Elementos de inferencia estadística: experimentos con uno y dos tratamientos

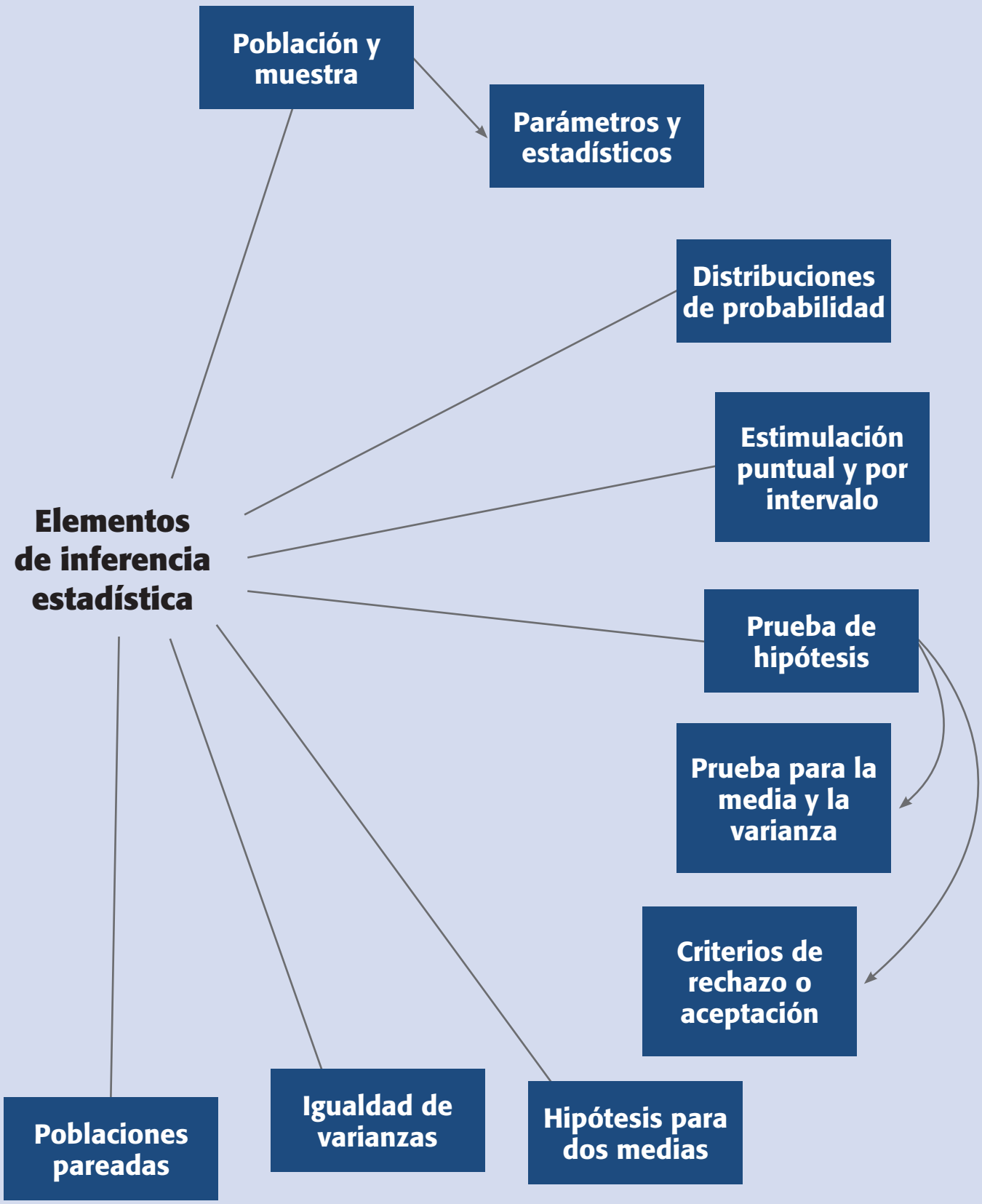
Sumario

- Población y muestra, parámetros y estadísticos
- Distribuciones de probabilidad e inferencia
- Estimación puntual y por intervalo
- Conceptos básicos de prueba de hipótesis
- Planteamiento de una hipótesis estadística
- Prueba para la media
- Prueba para la varianza
- Tres criterios de rechazo o aceptación equivalentes
- Hipótesis para dos medias: comparación de dos tratamientos
- Prueba para la igualdad de varianzas
- Poblaciones pareadas (comparación de dos medias con muestras dependientes)
- Resumen de fórmulas para procedimientos de prueba de hipótesis
- Uso de un software estadístico

Objetivos de aprendizaje

- Identificar los elementos de la inferencia estadística y su importancia en los diseños experimentales.
- Explicar el papel de las distribuciones de probabilidad en la inferencia estadística, así como la estimación puntual y por intervalo.
- Describir las pruebas para la media y la varianza, así como los conceptos básicos de prueba de hipótesis.
- Identificar las pruebas para la igualdad de varianzas.
- Distinguir las pruebas para comparar medias con muestras independientes y muestras pareadas.

Mapa conceptual



Conceptos clave

- Error tipo I
- Error tipo II
- Estadístico
- Estadístico de prueba
- Estimador puntual
- Grados de libertad
- Hipótesis estadística
- Inferencia estadística
- Intervalo de confianza
- Muestras pareadas
- Orden completamente al azar
- Potencia de la prueba
- Región de aceptación
- Región de rechazo
- Significancia observada
- Significancia predefinida

Población y muestra, parámetros y estadísticos

Una *población* o *universo* es una colección o totalidad de posibles individuos, especímenes, objetos o medidas de interés sobre los que se hace un estudio. Las poblaciones pueden ser finitas o infinitas. Si es *finita* y pequeña se pueden medir todos los individuos para tener un conocimiento "exacto" de las características (*parámetros*) de esa población. Por ejemplo, un parámetro que podría ser de interés es la proporción p de productos defectuosos, o la media, μ, de alguna variable medida a los productos. Si la población es *infinita* o grande es imposible e incosteable medir a todos los individuos, en este caso se tendrá que sacar una *muestra representativa* de dicha población, y con base en las características medidas en la muestra (*estadísticos*) se podrán hacer afirmaciones acerca de los parámetros de la población (figura 2.1).

Con frecuencia, las poblaciones de interés son los materiales, los productos terminados, partes o componentes, o algunos de los procesos. En muchos casos estas poblaciones se pueden suponer infinitas o grandes. Por ejemplo, en empresas con producción en masa no siempre es posible medir cada pieza de material que llega o las propiedades de cada producto terminado. Incluso, si la producción no es masiva, conviene imaginar al proceso como una población infinita o muy grande, debido a que el flujo del proceso no se detiene, es decir, no existe el último artículo producido mientras la empresa siga operando. En estos casos los procesos (poblaciones) se estudian mediante muestras de artículos extraídas en algún punto del proceso.

Un asunto importante será lograr que las muestras sean representativas, en el sentido de que tengan los aspectos clave que se desean analizar en la población. Una forma de lograr esa representatividad es diseñar de manera adecuada un muestreo aleatorio (azar), donde la selección no se haga con algún sesgo en una dirección que favorezca la inclusión de ciertos elementos en particular, sino que todos los elementos de la población tengan las mismas oportunidades de ser incluidos en la muestra. Existen varios métodos de muestreo aleatorio, por ejemplo: el simple, el estratificado, el muestreo sistemático y por conglomerados; cada uno de ellos logra muestras representativas en función de los objetivos del estudio y de ciertas circunstancias y características particulares de la población (véase Gutiérrez Pulido, 2005).

Población finita
Es aquella en la que se pueden medir todos los individuos para tener un conocimiento exacto de sus características.

Parámetros
Características que, mediante su valor numérico, describen a un conjunto de elementos o individuos.

Población infinita
Es aquella en la que la población es grande y es imposible e incosteable medir a todos los individuos.

Muestra representativa
Es una parte de una población, seleccionada adecuadamente, que conserva los aspectos clave de la población.

Inferencia estadística
Son las afirmaciones válidas acerca de la población o proceso basadas en la información contenida en la muestra.

Inferencia estadística

El objetivo de la *inferencia estadística* es hacer afirmaciones válidas acerca de la población o proceso con base en la información contenida en una muestra. Estas afirmaciones tienen por objetivo coadyuvar en la toma de decisiones. La inferencia estadística por lo general se divide en *estimación* y *prueba de hipótesis*, y se apoya en cantidades o datos estadísticos calculados a partir de las observaciones en la muestra. Un *estadístico* se define como cualquier función de los datos muestrales que no contiene parámetros desconocidos. Un ejemplo de estadístico es la media muestral $\bar{X}$ con la cual se tratan de hacer afirmaciones sobre la media, μ, que es un parámetro poblacional.

Un aspecto clave en la interpretación y utilización de cualquier estadístico es que se trata de una variable aleatoria, ya que su valor depende de los elementos que

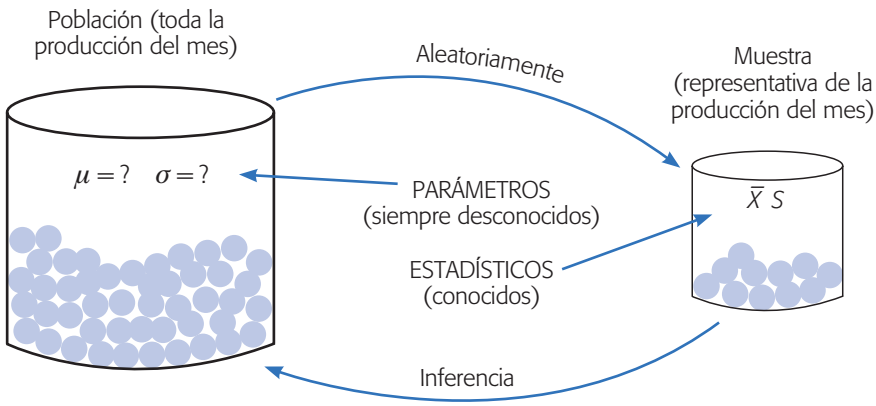


Figura 2.1 Relación entre población y muestra, parámetros y estadísticos.

son seleccionados en la muestra y, por lo tanto, varía de una muestra a otra. La forma de tomar en cuenta este hecho es conocer la distribución de probabilidad de cada estadístico. Como se verá más adelante, al conocer la distribución de probabilidad del estadístico se podrán hacer estimaciones acerca de cuál es el valor del parámetro poblacional, y también será posible probar o verificar la validez de hipótesis o conjeturas que se tengan sobre la población o proceso. Por ejemplo, un proveedor puede afirmar que el porcentaje de su producto que no cumple con especificaciones es de 0.1%, por lo que interesaría investigar, con base en una muestra, si esta afirmación se puede tomar como verdadera.

Estadístico
Cualquier función de los datos muestrales que no contiene parámetros desconocidos.

Distribuciones de probabilidad e inferencia

La *distribución de probabilidad* o *distribución de una variable aleatoria X* relaciona el conjunto de valores posibles de X (rango de X), con la probabilidad asociada a cada uno de estos valores y los representa a través de una tabla o por medio de una función planteada como una fórmula. Por ejemplo, sea la variable aleatoria dada por el estadístico media muestral, $\bar{X}$, entonces al conocer su distribución de probabilidad podremos saber cuáles son los valores que puede tomar $\bar{X}$ y cuáles son más probables.

Distribución de probabilidad de *X*
Relaciona el conjunto de valores de X con la probabilidad asociada con cada uno de estos valores.

En otras palabras, la distribución de probabilidad de la media muestral $\bar{X}$ señala qué valores se espera que tome $\bar{X}$, de acuerdo con los supuestos asumidos. De esta forma, la distribución de probabilidad hace que lo aleatorio no sea un capricho, y modela (describe, acota) los posibles valores de un estadístico muestral, con lo que al observar una realización específica de un estadístico se pueden corroborar o rechazar supuestos (prueba de hipótesis), o bien, hacer estimaciones poblacionales.

Las distribuciones de probabilidad que más se usan en intervalos de confianza y pruebas de hipótesis son las distribuciones: *normal*, *T de Student*, *ji-cuadrada* y *F*. En la figura 2.2 se representan las formas típicas de estas cuatro distribuciones. La distribución normal está completamente definida por sus parámetros, que son la media, μ, y la desviación estándar, σ. Por ejemplo, en la figura 2.2 se muestra la

distribución normal con $\mu = 0$ y $\sigma = 1$, que se simboliza con $N(0, 1)$ y se conoce como la *distribución normal estándar*.

En la figura 2.2 también se observa que, tanto la distribución normal estándar como la T de Student son simétricas y centradas en cero, mientras que las distribuciones ji-cuadrada y F son sesgadas y sólo toman valores positivos. Las cuatro distribuciones están relacionadas entre sí, ya que las distribuciones T de Student, ji-cuadrada y F se definen en términos de la distribución normal estándar. Los parámetros que definen por completo las distribuciones T de Student, ji-cuadrada y F, reciben el nombre de *grados de libertad*, que tienen que ver con los tamaños muestrales involucrados. Por ejemplo, si se tiene una muestra de tamaño 20, será de interés una distribución T de Student con 19 grados de libertad para hacer inferencia sobre la media poblacional; o una ji-cuadrada con 19 grados de libertad para hacer inferencias sobre la varianza poblacional.

Grados de libertad
Son parámetros que definen las distribuciones *T*, ji-cuadrada y *F*, y se determinan a partir de los tamaños muestrales involucrados.

La distribución T de Student tiende a la distribución normal estándar cuando el tamaño de muestra crece, y prácticamente es la misma distribución para $n > 45$. La diferencia básica entre las dos distribuciones es que la T de Student es más ancha (respecto del eje horizontal) en las colas (véase figura 2.2). La distribución normal estándar es una curva única, por ello existen tablas que proporcionan cualquier área o probabilidad de interés bajo esta curva. No pasa lo mismo con las otras distribucio-

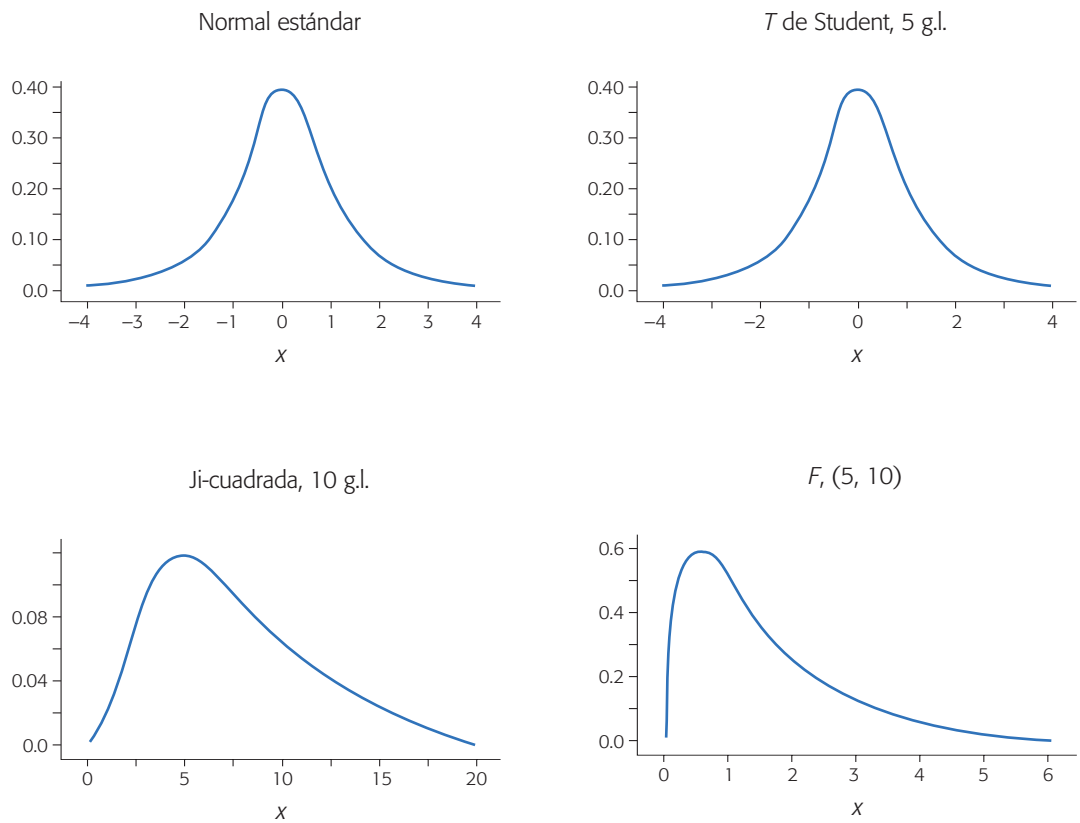


Figura 2.2 Muestra de las distribuciones de probabilidad de mayor uso en inferencia.

nes a las que hemos hecho referencia, ya que para cada tamaño muestral es una curva diferente. Por eso, las tablas de estas distribuciones sólo reportan los valores que separan las áreas de mayor uso en inferencia estadística (ver apéndice 2). En la actualidad es mejor utilizar un paquete estadístico para encontrar cualquier área o percentil que se quiera de cada distribución.

Como se muestra más adelante, las distribuciones normal y T de Student sirven para hacer inferencias sobre las medias; mientras que la distribución ji-cuadrada será de utilidad para hacer inferencias sobre varianzas y la distribución F se empleará para comparar varianzas. Es por esto que la distribución F es la de mayor relevancia en diseño de experimentos, dado que el análisis de la variabilidad que se observó en un experimento se hace comparando varianzas.

Uso de Excel

Se puede utilizar la hoja de cálculo de Excel (o algo equivalente) para calcular las probabilidades con la distribución normal. Para ello se utiliza la siguiente función:

$$DISTR.NORM(x, media, desv\_estándar, acum)$$

donde en la celda x se da el valor de referencia para el cálculo de probabilidades ($P(X \leq x)$), en *media* se da el valor de la media, μ, de la distribución normal con la que se quiere obtener probabilidades, y en *desv_estándar* se declara el valor de la desviación estándar, σ, de la distribución normal. Por último, *acum* es un valor lógico que determina la forma de la función, si el argumento *acum* es VERDADERO (se da un 1), la función *DISTR.NORM* devuelve la función de distribución acumulada ($P(X \leq x)$); si es FALSO (se da un 0), devuelve la función de densidad de probabilidad, es decir, da $f(x)$.

Estimación puntual y por intervalo

Las distribuciones de probabilidad que tienen una variable que representa cierta característica de una población se definen completamente cuando se conocen sus parámetros, pero cuando éstos no se conocen, será necesario estimarlos con base en los datos muestrales para hacer inferencias sobre la población. Por ejemplo, los parámetros de una distribución normal son la media, μ, y la desviación estándar, σ, que en caso de desconocerse será necesario estimarlos a partir de los datos en la muestra. Hay dos tipos de estimación: puntual y por intervalo.

Estimación puntual

Estimador puntual
Estadístico que estima un valor específico de un parámetro.

Un *estimador puntual* de un parámetro desconocido es un estadístico que genera un valor numérico simple, que se utiliza para hacer una estimación del valor del parámetro desconocido; por ejemplo, tres parámetros sobre los que con frecuencia se desea hacer inferencia son:

- La media μ del proceso (población).
- La varianza σ^2 o la desviación estándar σ del proceso.
- La proporción p de artículos defectuosos.

Los estimadores puntuales (estadísticos) más recomendados para estimar estos parámetros son, respectivamente:

- La media muestral $\hat{\mu} = \bar{X}$.
- La varianza muestral $\hat{\sigma}^2 = S^2$.
- La proporción de defectuosos en la muestra, $\hat{p} = x/n$, donde x es el número de artículos defectuosos en una muestra de tamaño n.

Por ejemplo, para estimar el grosor promedio de los discos producidos por un proceso, durante una semana se toma una muestra de $n = 125$ discos, y se obtiene que la media muestral es $\bar{X} = 1.179$. Este valor puede usarse como una estimación puntual de μ (la media del proceso).

Colocar un gorro (símbolo ˆ) sobre un parámetro es una manera general de denotar un estimador puntual del correspondiente parámetro, puesto que los estimadores no son únicos. Por ejemplo, la estimación de la media, $\hat{\mu}$, podría hacerse con el uso de la media muestral $\bar{X}$, la mediana X, o la moda, dado que las tres son diferentes medidas de la tendencia central de unos datos.

Estimación por intervalo

La estimación puntual de un parámetro se genera a través de un estadístico, y como el valor de éste es aleatorio porque depende de los elementos que fueron seleccionados en la muestra, entonces la estimación que se hace sobre el parámetro dependerá y variará de una muestra a otra. De esta forma, cuando se quiere tener mayor certidumbre sobre el verdadero valor del parámetro poblacional, será necesario obtener la información sobre qué tan precisa es la estimación puntual. Así, la estimación puntual dirá poco sobre el parámetro cuando la variación entre una estimación y otra es muy grande. Una forma de saber qué tan variable es el estimador, consiste en calcular la desviación estándar o *error estándar* del estadístico, visto como una variable aleatoria. Por ejemplo, consideremos la desviación estándar S y la media $\bar{X}$ de una muestra de tamaño n. Puesto que $\bar{X}$ es una variable aleatoria, ésta tiene su propia desviación o error estándar, que se puede estimar mediante $\hat{\sigma}_{\bar{X}} = S/\sqrt{n}$.

Una forma operativa de saber qué tan precisa es la estimación consiste en calcular un *intervalo de confianza* que indique un rango "donde puede estar el parámetro" con cierto nivel de seguridad o confianza. Construir un intervalo al $100(1 - \alpha)\%$ de confianza para un parámetro desconocido θ, consiste en estimar dos números (estadísticos) L y U, de manera que la probabilidad de que θ se encuentre entre ellos sea $1 - \alpha$, es decir,

$$P(L \leq \theta \leq U) = 1 - \alpha \qquad \textbf{(2.1)}$$

donde L y U forman el intervalo de confianza buscado $[L, U]$. La correcta interpretación de un intervalo de confianza es como sigue: si se obtuvieran 100 muestras inde-

pendientes de la misma población o proceso, cada una de tamaño n y para cada muestra se calculará el intervalo de confianza a 95% para el mismo parámetro, entonces se espera que 95 de los 100 intervalos contengan el verdadero valor de dicho parámetro. En la práctica se obtiene sólo un intervalo y se dice que el intervalo $[L, U]$ tiene una confianza de $100(1 - \alpha)\%$; esto tiene una interpretación constante, en el sentido de que el parámetro estará en el intervalo $100(1 - \alpha)\%$ de las veces que apliquemos el procedimiento.

Intervalo de confianza
Rango donde se estima que está el valor de un parámetro poblacional.

La *longitud del intervalo de confianza* es una medida de la precisión de la estimación. De aquí que es deseable que la longitud de los intervalos sea pequeña y con alto nivel de confianza. El ancho de los intervalos es mayor a medida que sea mayor la varianza de la población y el nivel de confianza exigido. El ancho del intervalo es menor si se incrementa el tamaño de la muestra.

Intervalo de confianza para una media

Por definición de *intervalo de confianza* se trata de encontrar dos números L y U, tales que el parámetro μ se encuentre entre ellos con una probabilidad de $1 - \alpha$. Esto es,

$$P(L \leq \mu \leq U) = 1 - \alpha$$

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de una población, con una distribución normal con media μ y varianza σ^2, ambas desconocidas. El procedimiento general para deducir el intervalo consiste en partir de un estadístico que involucra al parámetro de interés y que tiene una distribución conocida. Tal estadístico es

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

el cual sigue una distribución T de Student con $n - 1$ grados de libertad. Por lo tanto, en la tabla de esta distribución o en su gráfica se pueden ubicar dos valores críticos $t_{a/2}$ y $-t_{a/2}$, tales que:

$$P\left(-t_{\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2}\right) = 1 - \alpha$$

De aquí, despejando hasta dejar sólo en medio de las desigualdades al parámetro de interés, se llega a que

$$P\left(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha \qquad \textbf{(2.2)}$$

En este sentido, $L = \bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}$ y $U = \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$ son los números buscados que definen un intervalo al $100(1 - \alpha)\%$ para la media desconocida μ. En la tabla de la distribución T de Student se observa que para una muestra mayor o igual a 30, el intervalo al $100(1 - \alpha)\%$ para la media μ es aproximadamente $\bar{X} \pm 2\frac{S}{\sqrt{n}}$, o sea, la media más menos 2 veces su error estándar.

Ejemplo 2.1

En un proceso de inyección de plástico una característica de calidad del producto (disco) es su grosor, el cual debe ser de 1.20 mm con una tolerancia de ±0.10 mm. Así, el grosor del disco debe estar dentro de la especificación inferior, $EI = 1.10$, y la superior, $ES = 1.30$, para considerar que el proceso de inyección fue satisfactorio. Para evaluar esta característica de calidad, durante una semana se hace un muestreo sistemático en una línea de producción, y se obtienen 25 muestras de tamaño 5 cada una. Por lo tanto, al final se tiene una muestra de $n = 125$ y se obtiene la media muestral, $\bar{X} = 1.179$ mm y la varianza, $S^2 = 0.00071$, por lo que la estimación del error estándar de la media es

$$\frac{S}{\sqrt{n}} = \frac{0.0266}{11.18} = 0.0024$$

Cuando $n \geq 45$, la distribución T de Student es prácticamente igual a la distribución normal estándar, por lo tanto, de la tabla de la distribución normal se obtiene que $t_{\alpha/2} \simeq z_{\alpha/2} = 1.96$ para $\alpha = 0.05$. De aquí que el intervalo al $100(1-\alpha)\%$ de confianza para la media μ del grosor de los discos está dado por

$$\bar{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}} = 1.179 \pm 1.96\left(\frac{0.0266}{11.18}\right) = 1.179 \pm \mathbf{0}.00466$$

Se puede afirmar entonces que con una confianza de 95%, la media μ de grosor de los discos se encuentra en el intervalo [1.174, 1.184]. En el cálculo anterior al valor de 0.00466 se le conoce como error de estimación, porque hasta en 0.00466 puede diferir el estimador puntual $\bar{X}$ del parámetro poblacional μ.

Tamaño de la muestra. En ocasiones es necesario calcular el tamaño de muestra n para lograr que la estimación de una media poblacional μ tenga como error máximo a un número E. En este caso, como el error de estimación está dado por $E = t_{(\alpha/2,\, n-1)}S/\sqrt{n}$, entonces despejando n obtenemos que

$$n = \frac{t^2_{(\alpha/2,\, n-1)}S^2}{E^2}$$

Como $t_{(\alpha/2,\, n-1)}$ depende de n y ésta es la incógnita, entonces para propósitos prácticos y con tamaños de muestra mayores que 30, el valor de $t_{(\alpha/2,\, n-1)}$ puede tomarse como 2. De esta manera,

$$n = \frac{4S^2}{E^2}$$

donde S^2 es un estimador de la varianza. Por ejemplo, si en el caso del grosor medio de los discos se quisiera un error máximo de $0.004 = E$, entonces se requiere,

$$n = \frac{4(0.00071)}{(0.004)^2} = 177.5 \approx 178\,.$$

Intervalo para la varianza

De manera similar a como se obtiene el intervalo para la media, es posible deducir intervalos de confianza para cualquier parámetro. En particular, para construir un intervalo de confianza para la varianza σ^2, la distribución de referencia es una ji-cuadrada con $n-1$ grados de libertad, ya que bajo el supuesto de que la variable de interés tiene una distribución normal con media y varianza desconocidas, el estadístico $(n-1)S^2/\sigma^2$ sigue la distribución ji-cuadrada con $n-1$ grados de libertad. De esta manera, con un poco de álgebra, se llega a que el intervalo de confianza para la varianza está dado por

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,\,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,\,n-1}} \qquad \textbf{(2.3)}$$

donde $\chi^2_{\alpha/2,n-1}$ y $\chi^2_{1-\alpha/2,n-1}$ son puntos críticos de la distribución ji-cuadrada con $n-1$ grados de libertad y se leen en la tabla de esta distribución para el valor de α dado. Es decir, $P(X > \chi^2_{\alpha/2,n-1}) = \alpha/2$.

Ejemplo 2.2

En el proceso de fabricación de discos para computadoras, una de las variables críticas es el rendimiento de formato. Se toma una muestra aleatoria de $n = 10$ discos de la producción del turno de la mañana. Se formatean y se reporta el rendimiento de cada disco. Los datos obtenidos son: 96.11, 91.06, 93.38, 88.52, 89.57, 92.63, 85.20, 91.41, 89.79, 92.62. Con base en estos datos interesa estimar puntualmente y por intervalo la media y la desviación estándar para la población de discos de dicho turno.

Los estimadores puntuales para la media y la desviación estándar resultan ser

$$\bar{X} = \frac{\sum_{i=1}^{10} X_i}{10} = 91.03 \quad \text{y} \quad S = \sqrt{\frac{\sum_{i=1}^{10}(X_i - \bar{X})^2}{9}} = 2.99$$

Suponiendo distribución normal, el intervalo al 95% de confianza para la media μ está dado por

$$\left[\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right] = \left[91.03 - 2.26\frac{2.99}{\sqrt{10}},\ 91.03 + 2.26\frac{2.99}{\sqrt{10}}\right] = [88.89,\ 93.17]$$

donde el valor del punto crítico $t_{\alpha/2} = t_{0.025} = 2.26$ se lee en las tablas para la distribución T de Student con 9 grados de libertad que se localiza en el apéndice. Con una confianza de 95% se espera que el rendimiento promedio de los discos producidos durante ese turno esté entre 88.89 y 93.17. El correspondiente intervalo para la desviación estándar σ se obtiene sacando la raíz cuadrada al intervalo para la varianza σ^2 dado en la relación (2.3). Así, el intervalo para σ está dado por

$$\left[\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2,\,n-1}}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,\,n-1}}}\right] = \left[\sqrt{\frac{(9)(2.99)^2}{19.02}}, \sqrt{\frac{(9)(2.99)^2}{2.70}}\right] = [2.05,\ 5.46]$$

donde los valores críticos $\chi^2_{\alpha/2,\,n-1} = \chi^2_{0.025,9} = 19.02$ y $\chi^2_{1-\alpha/2,\,n-1} = \chi^2_{0.975,9} = 270$ se obtienen de la tabla de la distribución ji-cuadrada, que está en el apéndice, o también se pueden consultar usando un software. Así, con una confianza de 95% se espera que la desviación estándar del rendimiento de los discos producidos durante ese turno esté entre 2.05 y 5.46.

Cuando no se está satisfecho con la amplitud del intervalo, entonces será necesario incrementar la precisión de la estimación, y esto se hace aumentando el tamaño de muestra.

Intervalo para la proporción

Bajo el supuesto de que el número de artículos defectuosos en una muestra sigue una distribución binomial, y suponiendo que se inspecciona una cantidad grande de n artículos y se encuentra una proporción $\hat{p}$ de defectuosos, se puede construir un intervalo de confianza para la proporción poblacional p, apoyándose en la aproximación de la distribución binomial por la normal. En estas condiciones se puede afirmar que la proporción muestral $\hat{p}$ sigue una distribución normal con media p y varianza $\frac{p(1-p)}{n}$. Con el uso de la misma argumentación que en el intervalo para la media, se deduce que el intervalo de confianza para la proporción es de la forma

$$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

donde $Z_{\alpha/2}$ es un percentil de tabla de la distribución normal estándar que está en el apéndice.

Tabla 2.1 Resumen de fórmulas para intervalos de confianza.

Parámetro	Límite inferior	Límite superior
μ	$\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}$	$\bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$
σ^2	$\frac{(n-1)S^2}{\chi^2_{\alpha/2,\,n-1}}$	$\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,\,n-1}}$
p	$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
$\mu_1 - \mu_2$	$(\bar{X}_1 - \bar{X}_2) - t_{n_1+n_2-2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ donde $S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$	$(\bar{X}_1 - \bar{X}_2) + t_{n_1+n_2-2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
$\frac{\sigma_1^2}{\sigma_2^2}$	$\frac{S_1^2}{S_2^2}F_{1-\alpha/2,\,n_2-1,\,n_1-1}$	$\frac{S_1^2}{S_2^2}F_{\alpha/2,\,n_2-1,\,n_1-1}$
$p_1 - p_2$	$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$(\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Ejemplo 2.3

Se quiere estimar la proporción p de artículos defectuosos en un lote de 2 000 (población). Para ello, se toma una muestra aleatoria de $n = 100$ artículos y se encuentra que de éstos, $x = 5$, son defectuosos. Por lo tanto, un estimador puntual de p es $\hat{p} = 5/100 = 0.050$. Si se quiere estimar p por intervalo, entonces de acuerdo con lo explicado antes, un intervalo al 95% de confianza está dado por

$$0.050 \pm 1.96\sqrt{\frac{0.05(1-0.05)}{100}} = 0.050 \pm 0.043$$

de aquí que, con una confianza de 95%, p está entre 0.007 y 0.093, en términos porcentuales entre 0.7% y 9.3%. En el cálculo anterior, al valor de 0.043 se le conoce como *error de estimación*, porque hasta en ese valor puede diferir $\hat{p}$ de p.

Tamaño de muestra. Si se quiere estimar el tamaño de la muestra n, que es necesario para estimar p con un error máximo de E, entonces dado que $E = Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$; si despejamos de aquí a n obtenemos que

$$n = \frac{Z^2_{\alpha/2}\hat{p}(1-\hat{p})}{E^2}$$

donde $\hat{p}$ es una estimación del valor de p. Por ejemplo, si en el problema anterior se quisiera un error máximo de $E = 0.03$, con una confianza de 95%, entonces se requiere que $n = (1.96)^2\,(0.05)(1 - 0.05)/(0.03)^2 \approx 203$. En ocasiones, cuando no se sabe nada de p en la fórmula anterior, se supone $\hat{p} = 0.5$.

Resumen de fórmulas para intervalos de confianza

En la tabla 2.1 se muestran las fórmulas para calcular los intervalos de confianza más usuales. Además de los intervalos para un parámetro ya presentados, en la tabla se incluyen las fórmulas correspondientes para intervalos de confianza que involucran a dos parámetros, como son: diferencias de medias, diferencias de proporciones y cocientes de varianzas. Estos intervalos proveen información sobre la igualdad estadística de los parámetros correspondientes a las dos poblaciones de interés. Note que los cálculos involucran a los estimadores puntuales obtenidos con cada muestra. En la tabla, la notación $z_{1-\alpha/2}$, $t_{1-\alpha/2}$, $\chi^2_{1-\alpha/2,\,n-1}$ y $F_{1-\alpha/2,\,n_2-1,\,n_1-1}$, se refiere a puntos críticos de la correspondiente distribución. Estos valores se determinan fácilmente con el uso de un software estadístico o de las tablas dadas en el apéndice.

Conceptos básicos de prueba de hipótesis

Un estudio experimental o una investigación, por lo general tiene como último objetivo, responder en forma segura ciertas preguntas y/o tomar decisiones. En este contexto, el experimentador tiene *a priori* ciertas creencias o hipótesis que desea comprobar. Por ejemplo:

- Los tres proveedores del material x tienen el mismo nivel de calidad.
- El porcentaje de este ingrediente afecta el resultado de la mezcla.

- El tiempo de espera de esta operación es de tres horas, en promedio.
- Si aumentamos la cantidad de reactivo se elimina el problema.

A continuación se describen los conceptos básicos de prueba de hipótesis, es decir, los pasos fundamentales de cualquier procedimiento de prueba de hipótesis, como son: planteamiento de la hipótesis, estadístico de prueba y criterio de rechazo.

Planteamiento de una hipótesis estadística

Hipótesis estadística
Es una afirmación sobre los valores de los parámetros de una población o proceso, que puede probarse a partir de la información contenida en una muestra.

Una *hipótesis estadística* es una afirmación sobre los valores de los parámetros de una población o proceso, que es susceptible de probarse a partir de la información contenida en una muestra representativa que es obtenida de la población. Por ejemplo, la afirmación "este proceso produce menos de 8% de defectuosos" se puede plantear estadísticamente, en términos de la proporción p desconocida de artículos defectuosos que genera el proceso, como se hace a continuación.

$$H_0 : p = 0.08 \text{ (la proporción de defectuosos es 0.08)}$$
$$H_A : p < 0.08 \text{ (la proporción es menor a 0.08)} \quad \textbf{(2.4)}$$

A la expresión $H_0 : p = 0.08$ se le conoce como *hipótesis nula* y $H_A : p < 0.08$ se le llama *hipótesis alternativa*. El nombre de hipótesis nula se deriva del hecho de que comúnmente se plantea como una igualdad, lo cual facilita el tener una distribución de probabilidad de referencia específica. En general, la estrategia a seguir para probar una hipótesis es suponer que la hipótesis nula es verdadera, y que en caso de ser rechazada por la evidencia que aportan los datos, se estará aceptando la hipótesis alternativa. Así, en el caso de las proporciones, la afirmación que se desea probar se aceptará como cierta, sólo en caso de rechazar la hipótesis nula.

Supongamos ahora que la afirmación a probar es "este proceso produce 8% de defectuosos". Observe que la afirmación señala que su falsedad se da, tanto si se observan menos de 8% de defectuosos como si se observan más de 8% de defectuosos. En este sentido, el planteamiento estadístico debe ser:

$$H_0 : p = 0.08 \text{ (la proporción de defectuosos es 0.08)}$$
$$H_A : p \neq 0.08 \text{ (la proporción es diferente a 0.08)} \quad \textbf{(2.5)}$$

Ahora, lo que se desea concluir es la hipótesis nula. Nótese la diferencia entre las hipótesis alternativas en las expresiones (2.4) y (2.5). En (2.4) H_A se conoce como *hipótesis alternativa de un solo lado* (*unilateral*), ya que la única manera de rechazar H_0 es teniendo valores de la proporción muestral $\hat{p}$ significativamente más pequeños que 0.08. Asimismo, en (2.5) H_A se llama *hipótesis alternativa de dos lados* (*bilateral*), ya que la evidencia en contra de H_0 se obtiene con valores pequeños o grandes de la proporción muestral $\hat{p}$. Así, la elección de la hipótesis alternativa en cuanto a si debe ser unilateral o bilateral depende de la afirmación que se quiera probar.

Otro aspecto importante es la selección del valor del parámetro que especifica la hipótesis nula, esto es, ¿por qué 0.08 en las hipótesis de las expresiones (2.4) y (2.5)? Este valor se elige de manera que separe dos situaciones que llevan a tomar diferentes acciones. Por ejemplo, en la hipótesis dada en (2.4) se eligió 0.08, porque ésta es la proporción de defectuosos reportada el mes anterior, y después de implementar un programa de mejora se quiere ver si dio el resultado esperado. En caso de no rechazar H_0 se concluiría que el programa no funcionó y que se deben tomar medidas adicionales para bajar la proporción de defectuosos.

Estadístico de prueba

Probar una hipótesis consiste en investigar si lo afirmado por la hipótesis nula es verdad o no. La estrategia de prueba parte del supuesto de que H_0 es verdadera, y si los resultados de la investigación contradicen en forma suficiente dicho supuesto, entonces se rechaza H_0 y se acepta la hipótesis alternativa. En caso de que los resultados de la investigación no demuestren claramente la falsedad de H_0, ésta no se rechaza. Es decir, *la hipótesis nula es verdadera mientras no se demuestre lo contrario*.

Una vez planteada la hipótesis, se toma una muestra aleatoria de la población de estudio o se obtienen datos mediante un experimento planeado de acuerdo con la hipótesis. El *estadístico de prueba* es un número calculado a partir de los datos y la hipótesis nula, cuya magnitud permite discernir si se rechaza o no la hipótesis nula H_0. Al conjunto de posibles valores del estadístico de prueba que llevan a rechazar H_0, se le llama *región o intervalo de rechazo* para la prueba, y a los posibles valores donde *no* se rechaza H_0 se les llama *región o intervalo de aceptación*. Por ejemplo, para las hipótesis planteadas en (2.4) y (2.5), el estadístico de prueba está dado por

$$z_0 = \frac{\hat{p} - 0.08}{\sqrt{0.08(1-0.08)/n}} \qquad \textbf{(2.6)}$$

donde $\hat{p}$ es la proporción de defectuosos que se encontró en una muestra de n artículos inspeccionados. Si H_0 es verdad, el estadístico z_0 sigue aproximadamente la distribución normal estándar; la aproximación es mejor mientras más grande es el valor de n. En general, se requiere $np > 10$ para una buena aproximación; en este caso, con $np \geq 120$ unidades inspeccionadas sería suficiente.

Por ejemplo, supongamos que se toma una muestra de $n = 150$ piezas y de ellas $x = 20$ son defectuosas, entonces el valor de la proporción es $\hat{p} = x/n = 0.13$. Vamos a ver si esto implica una diferencia suficiente para rechazar que $p = 0.08$. Por lo pronto, el valor estadístico es $z_0 = 2.41$.

Estadístico de prueba
Número calculado a partir de los datos y de H_0, cuya magnitud permite discernir si se rechaza o no la hipótesis nula.

Región de rechazo
Es el conjunto de posibles valores del estadístico de prueba que llevan a rechazar la hipótesis nula.

Región de aceptación
Son los posibles valores del estadístico de prueba donde no se rechaza la hipótesis nula.

Criterio de rechazo

El estadístico de prueba, construido bajo el supuesto de que H_0 es verdad, es una variable aleatoria con distribución conocida. Si efectivamente H_0 es verdad, el valor del estadístico de prueba debería caer dentro del rango de valores más probables de su distribución asociada, el cual se conoce como *región de aceptación*. Si cae en una de las colas de su distribución asociada, fuera del rango de valores más probables (en

la *región de rechazo*), es evidencia en contra de que este valor pertenece a dicha distribución (véase figura 2.3). De aquí se deduce que debe estar mal el supuesto bajo el cual se construyó, es decir, H_0 debe ser falsa.

Pruebas de una y dos colas (unilaterales y bilaterales). La ubicación de la región o intervalo de rechazo depende de si la hipótesis es *bilateral* o *unilateral*. Como se vio en el caso de las proporciones, una hipótesis es bilateral cuando la hipótesis alternativa (H_A) es del tipo "no es igual" ($\neq$); y es unilateral cuando la alternativa es del tipo "mayor que" ($>$) o "menor que" ($<$). Cuando es bilateral, como en la expresión (2.5), la región de rechazo está repartida de manera equitativa entre ambas colas de la distribución del estadístico de prueba. Pero si la hipótesis es unilateral, como en la expresión (2.4), la evidencia en contra de la hipótesis nula se ubica en un solo lado de la distribución, esto es, la región de rechazo sólo se concentra en una de las colas. En la expresión (2.4) la región de rechazo se concentra en el lado izquierdo de la distribución del estadístico dado por (2.6) (véase figura 2.3).

Para probar la hipótesis sobre la proporción se calcula el estadístico de prueba z_0 de la ecuación (2.6) y se verifica si cae en la región de rechazo o aceptación. Por ejemplo, si las hipótesis planteadas son las expresiones (2.4) se rechaza H_0 si $z_0 < -z_a$; si las hipótesis están dadas por las relaciones (2.5) se rechaza H_0 si $z_0 < -z_{a/2}$ o si $z_0 > z_{a/2}$, o simplemente, si $|z_0| > z_{a/2}$. En la figura 2.3 esto equivale a que z_0 caiga en el rango de las áreas sombreadas, de acuerdo con la hipótesis de que se trate.

Si queremos probar la hipótesis bilateral con una confianza de 95%, entonces $z_{\alpha/2} = 1.96$; además, como $\hat{p} = 0.13$ y $z_0 = 2.41$, entonces $z_0 > 1.96$; por lo tanto, se rechaza $H_0 : p = 0.08$. De alguna forma, esto ya se intuía, puesto que la proporción muestral había sido $\hat{p} = 0.13$.

Si en lugar de tener $x = 20$ defectos, se tuvieran $x = 15$, entonces $\hat{p} = 0.10$. Al sustituir esto en (2.6) con $n = 150$, se obtiene que $z_0 = 0.90$ que no es mayor que $Z_{\alpha/2} = 1.96$. De aquí que no se rechace $H_0 : p = 0.08$. Es decir, en este caso $\hat{p} = 0.10$ no es evidencia suficiente contra $H_0 : p = 0.08$.

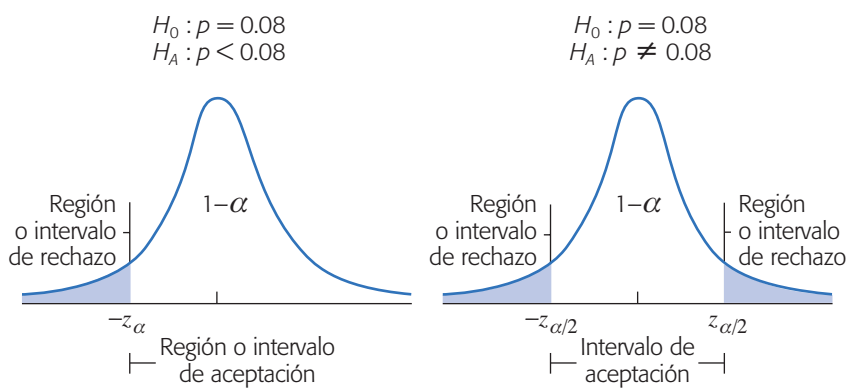


Figura 2.3 Hipótesis unilateral y bilateral, regiones de aceptación y rechazo.

El riesgo de una decisión equivocada: errores tipo I y tipo II

Probar una hipótesis estadística es una decisión probabilística, por lo que existe el riesgo de cometer un *error tipo I* o un *error tipo II*. El primero ocurre cuando se rechaza H_0 cuando ésta es verdadera, y el error tipo II es cuando se acepta H_0 y ésta es falsa. En toda prueba de hipótesis cada tipo de error tiene una probabilidad de ocurrir. Con α y β se denotan las probabilidades de los errores tipo I y II, respectivamente. Así,

$\alpha = P\{\text{error tipo I}\}$ = probabilidad de rechazar H_0 siendo verdadera
$\beta = P\{\text{error tipo II}\}$ = probabilidad de aceptar H_0 siendo falsa

Error tipo I
Es cuando se rechaza una H_0 que es verdadera.

Error tipo II
Es cuando se acepta una H_0 que es falsa.

Potencia de la prueba
Es la probabilidad de rechazar H_0 cuando es falsa.

A $1 - \beta$ se le llama *potencia de la prueba*, y es la probabilidad de rechazar H_0 cuando es falsa. A α también se le conoce como la *significancia dada de la prueba* y es la probabilidad de la región o intervalo de rechazo; su valor se especifica por parte del investigador desde que planea el estudio. Por lo general se utilizan los valores $\alpha = 0.05$ o 0.01, dependiendo del riesgo que se quiera admitir en la conclusión. Mientras más pequeño es el valor de α se requiere más evidencia en los datos para rechazar H_0. Por ejemplo, si la acción a tomar después de rechazar H_0 implica una inversión fuerte de recursos, se recomienda utilizar $\alpha = 0.01$ para tener mayor confianza de que la decisión será la adecuada. Si la decisión no implica una inversión fuerte, es suficiente trabajar con $\alpha = 0.05$, que es el valor más utilizado para este riesgo. Esto es, un valor más pequeño que α no necesariamente será mejor, ya que si se admite poco riesgo ($\alpha \leq 0.01$) se está truncando la posibilidad de muchos cambios que serían positivos para la empresa. Utilizar $\alpha = 0.05$ significa que por cada 100 veces independientes que se aplica el procedimiento y se rechaza H_0, se espera que en un promedio de 95 veces, tal decisión sea la correcta.

Por lo general, en las pruebas de hipótesis se especifica el valor de α y se diseña la prueba de tal forma que el valor de β sea pequeño. Esto es, la probabilidad del *error tipo I* se controla directamente, mientras que la probabilidad de *error tipo II* se controla de manera indirecta con el tamaño de la muestra, ya que a más datos β será menor. En otras palabras, con una muestra grande es mayor la *potencia de la prueba*,[1] es decir, se incrementa la probabilidad de rechazar H_0 si ésta es falsa.

En la práctica suele ser más delicado cometer el error tipo I que el error tipo II, debido a que en la mayoría de las hipótesis el rechazar H_0 implica objetar algo que se acepta de manera convencional. No rechazar H_0 implica, en muchos casos, seguir como hasta ahora. Por lo anterior, es común que se controle sólo el error tipo I, mientras que el error tipo II se deja libre como si su magnitud no importara.

Lo cierto es que el error tipo II también importa y la magnitud de su probabilidad debe ser pequeña (se recomienda $\beta = 0.10$). El problema es que controlar a β

[1] Es posible afirmar que, en general, es deseable que una prueba estadística sea potente. Sin embargo, cuando el tamaño de la muestra se incrementa en exceso (a tamaños en cientos) se llega a tener una potencia excesiva, que lleva al extremo de rechazar H_0 cuando es verdadera desde el punto de vista práctico.

tiene varios problemas; por ejemplo, muchas veces se requieren grandes tamaños muestrales o se deben realizar muchas repeticiones en el experimento. Por ello, en este libro no enfatizamos el control del error tipo II, pero damos las recomendaciones del número de repeticiones que deben obtenerse en cada experimento para tener un valor pequeño de β.

Prueba para la media

Cuando se estudia el comportamiento de un proceso o un fenómeno suelen interesar su media y varianza (o desviación estándar). En particular, al estudiar la media μ, es de interés preguntarse si ésta es igual, mayor o menor a cierto valor μ_0, donde μ_0 es un número conocido. Por ejemplo, puede ser de interés investigar si el rendimiento promedio del proceso durante esta semana es igual, mayor o menor que el de la semana anterior, μ_0. Cualquiera de estas tres preguntas se responden planteando una hipótesis estadística adecuada.

Las hipótesis se pueden probar suponiendo la varianza poblacional σ^2 conocida o desconocida. Sin embargo, como en la mayoría de los problemas es irreal suponer de antemano que se conoce la varianza, nos limitamos a describir el caso cuando σ^2 no se conoce.

Prueba para la media con varianza desconocida

Sea X una variable aleatoria con distribución normal con media μ y varianza σ^2, ambas desconocidas. Se quiere probar la hipótesis de que la media es igual a cierto valor μ_0. Es decir, la hipótesis a probar es

$$H_0 : \mu = \mu_0$$
$$H_A : \mu \neq \mu_0 \qquad \textbf{(2.7)}$$

Para probar esta hipótesis se toma una muestra aleatoria de tamaño n de los posibles valores de la variable X y se calcula el estadístico de prueba:

$$t_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \qquad \textbf{(2.8)}$$

donde S es la desviación estándar de los datos. Bajo el supuesto de que H_0 es verdadera, este estadístico se distribuye T de Student con $n - 1$ grados de libertad. Se rechaza H_0 si el valor absoluto del estadístico de prueba es mayor que el valor crítico de la distribución, es decir, se rechaza H_0 si $|t_0| > t_{\alpha/2}$. Recordemos que $t_{\alpha/2}$ es el punto crítico de la distribución T de Student, tal que $P(t > t_{\alpha/2}) = \alpha/2$; o sea, las áreas bajo la curva a la derecha del punto $t_{\alpha/2}$ y a la izquierda de $-t_{\alpha/2}$ son iguales a $\alpha/2$ (véase figura 2.4). Estos valores críticos se obtienen de la tabla de la distribución T de Student dada en el apéndice.

Una breve justificación del criterio de rechazo para la prueba anterior es la siguiente: por teoría estadística se sabe que bajo el supuesto de que $H_0 : \mu = \mu_0$ es verdadera, el estadístico de prueba t_0 se distribuye T de Student con $n - 1$ grados de libertad y, en consecuencia, hay una probabilidad de $100(1 - \alpha)\%$ de que el valor

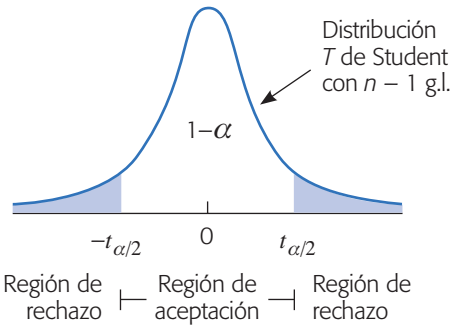


Figura 2.4 Regiones de rechazo y de aceptación para hipótesis (2.7).

de t_0 caiga entre $-t_{\alpha/2}$ y $t_{\alpha/2}$. De aquí que, si la muestra produce un valor de t_0 fuera de estos límites, entonces tal valor de t_0 es evidencia de que H_0 es falsa. Por el contrario, si t_0 cae entre $-t_{\alpha/2}$ y $t_{\alpha/2}$ es evidencia a favor de la veracidad de H_0 ya que no existe ninguna contradicción. Obsérvese que la región de rechazo dada por la unión de intervalos $(-\infty, -t_{\alpha/2}) \cup (-t_{\alpha/2}\ \infty)$ está determinada por la probabilidad α del error tipo I (véase figura 2.4).

En aquellas situaciones en que se desea rechazar $H_0 : \mu = \mu_0$ sólo cuando $\mu > \mu_0$, la hipótesis alternativa es unilateral:

$$H_0 : \mu = \mu_0$$
$$H_A : \mu > \mu_0 \qquad \textbf{(2.9)}$$

En este caso se rechaza H_0 si $t_0 > t_\alpha$. Por otra parte, si lo que interesa es rechazar $H_0 : \mu = \mu_0$ sólo cuando $\mu < \mu_0$, entonces ahora, la hipótesis unilateral se plantea de la forma:

$$H_0 : \mu = \mu_0$$
$$H_A : \mu < \mu_0 \qquad \textbf{(2.10)}$$

y se rechaza H_0 si $t_0 < -t_\alpha$.

Ejemplo 2.4

Peso de costales. Un fabricante de dulces compra costales de azúcar a cierto ingenio. Según los vendedores, los costales tienen un peso medio de 50.1 kg, con una varianza de ($\sigma^2 = 0.5$). El comprador sospecha que el peso medio es menor. Para confirmar su sospecha decide contrastar las hipótesis:

$$H_0 : \mu = 50.1$$
$$H_A : \mu < 50.1 \qquad \textbf{(2.11)}$$

con un nivel de significancia de 5% ($\alpha = 0.05$). Para ello, selecciona de manera aleatoria tres bultos de cada uno de los siguientes cinco pedidos. Pesa los 15 bultos y obtiene que $\bar{X} = 49.4$ y $S^2 = 1.2$. De esta manera, el estadístico de prueba calculado de acuerdo con la expresión (2.8), está dado por,

$$t_0 = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} = \frac{\sqrt{15}(49.4 - 50.1)}{\sqrt{1.2}} = -2.47$$

De las tablas de la distribución T de Student con $n - 1 = 14$ grados de libertad, para $\alpha = 0.05$, se lee el valor crítico $t_{0.05,\,14} = 1.76$. Como $t_0 = -2.47 < -1.76 = -t_{0.05,\,14}$, se rechaza la hipótesis H_0 (figura 2.5). Es decir, se rechaza la afirmación del vendedor de que los bultos tienen un peso medio de 50.1, y además la evidencia señala que dicho peso es menor que el declarado.

Prueba para la varianza

En el ejemplo 2.4 sobre el peso de costales, a simple vista se puede notar que la varianza $\sigma^2 = 0.5$, declarada por el vendedor, es bastante diferente que la varianza muestral $S^2 = 1.2$, lo cual lleva a sospechar que su afirmación sobre la varianza del proceso es falsa. El hecho de que los dos números sean distintos no significa que sean estadísticamente diferentes, de aquí la necesidad de contrastar o probar las hipótesis:

$$H_0 : \sigma^2 = 0.5$$
$$H_A : \sigma^2 > 0.5$$

y de esta manera comprobar si esa diferencia es estadísticamente significativa. Esta hipótesis es un caso particular de la siguiente:

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_A : \sigma^2 > \sigma_0^2$$

donde σ_0^2 es un valor conocido (0.5 en el ejemplo). Para probar esta hipótesis y bajo el supuesto de distribución normal, se utiliza el siguiente estadístico de prueba

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

donde n es el tamaño de la muestra. Si H_0 es verdadera χ_0^2 sigue una distribución ji-cuadrada con $n - 1$ grados de libertad. Por ello, se rechaza H_0 si $\chi_0^2 > \chi_\alpha^2$, donde χ_α^2 es un punto crítico que se obtiene de la tabla de distribución ji-cuadrada. Si aplicamos lo anterior al caso de la varianza del peso de los costales, obtenemos que

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{14 \times 1.2}{0.5} = 33.6$$

el cual, bajo el supuesto de normalidad, sigue una distribución ji-cuadrada con 14 grados de libertad cuando H_0 es verdadera. En la tabla de distribución ji-cuadrada se lee que χ_α^2, con $\alpha = 0.05$ y 14 grados de libertad es igual a 23.68. Como $\chi_0^2 = 33.6 >$

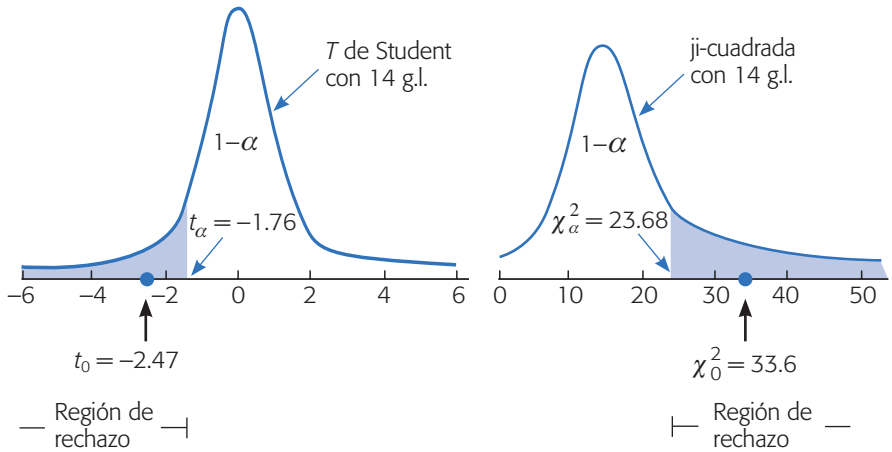


Figura 2.5 Resultados de las hipótesis para la media y para la varianza del peso de costales con $\alpha = 0.05$.

$23.68 = \chi^2_\alpha$ se rechaza H_0 y se acepta la hipótesis unilateral H_A (véase figura 2.5). Es decir, la varianza reportada por el vendedor para el peso de los costales es falsa y, en realidad, la variabilidad del peso de los costales es mayor.

Tanto el estadístico t_0 de la hipótesis sobre la media, como el estadístico χ^2_0 de la hipótesis sobre la varianza, cayeron en las respectivas regiones de rechazo, lo cual se representa en la figura 2.5.

Si la hipótesis alternativa para la varianza es bilateral, entonces se rechaza H_0 si $\chi^2_0 > \chi^2_{\alpha/2}$ o si $\chi^2_0 < \chi^2_{1-\alpha/2}$.

Tres criterios de rechazo o aceptación equivalentes

Al menos en las hipótesis más usuales, existen tres criterios equivalentes para decidir cuándo rechazar la hipótesis nula y, en consecuencia, aceptar la hipótesis alternativa. La equivalencia es en el sentido de que los tres llevan invariablemente a la misma decisión en términos de rechazar o no a H_0. Sin embargo, algunos de estos métodos proporcionan información adicional sobre la decisión que se está tomando, por lo que en algunas situaciones puede resultar ventajoso usar un criterio y no otro.

Estadístico de prueba frente a valor crítico

Éste es el criterio que utilizamos en el ejemplo previo y es el que tradicionalmente se empleaba antes de las facilidades que ahora provee la computación; por ello, es el que se explica en muchos libros de texto. Este método consiste en *rechazar H_0 si el estadístico de prueba cae en la región de rechazo que está delimitada por el valor crítico*. Debe tenerse cuidado de comparar los valores adecuados, dependiendo de la hipótesis alternativa de que se trata. Cuando los cálculos se hacen de forma manual, este criterio es el único que comúnmente se usa. No obstante, este método tradicional es el que da menos información adicional acerca de la decisión tomada.

Significancia observada frente a significancia predefinida

Significancia predefinida
Es el riesgo máximo que el experimentador está dispuesto a correr con respecto al error tipo I.

Significancia observada
Es el área bajo la distribución de referencia más allá del valor del estadístico de prueba. Se conoce como valor-*p*.

La *significancia predefinida* que se denota con α, es el riesgo máximo que el experimentador está dispuesto a correr por rechazar H_0 indebidamente (error tipo I). Mientras que la *significancia observada* o *calculada*, también conocida como *p-value* o valor-*p*, es el área bajo la distribución de referencia más allá del valor del estadístico de prueba. La expresión "más allá del estadístico de prueba" significa, por ejemplo en la prueba *T* bilateral, el área bajo la curva fuera del intervalo $[-t_0, t_0]$, es decir:

$$\text{valor-}p = P(T < -t_0) + P(T > +t_0)$$

donde *T* es una variable que tiene una distribución *T* de Student con $n - 1$ grados de libertad. Si la prueba es unilateral de cola derecha (izquierda), la significancia observada es el área bajo la curva de la distribución a la derecha (izquierda) de t_0. De lo anterior se desprende que H_0 se rechaza si la *significancia observada es menor que la significancia dada*, o sea, si valor-$p < \alpha$.

Este criterio es mejor que el anterior porque la significancia observada se puede ver como la probabilidad o evidencia a favor de H_0, por lo tanto, representa una medida de la contundencia con la que se rechaza o no la hipótesis nula. Por ejemplo, si la significancia observada o valor-*p* es igual a 0.0001, entonces sólo hay una probabilidad a favor de H_0 de 0.0001, por lo que se rechazaría la hipótesis nula con un riesgo tipo I de 0.0001, que es menor del que se está dispuesto a admitir, típicamente $\alpha = 0.05$. En otras palabras, un valor-$p = 0.0001$ nos señala que el valor observado del estadístico de prueba prácticamente no tenía ninguna posibilidad de ocurrir si la hipótesis nula fuera verdadera, lo que lleva a concluir de manera contundente que la hipótesis nula debe rechazarse.

En la figura 2.6 se muestra, utilizando una hipótesis bilateral, que cuando ocurre el evento $|t_0| < t_{\alpha/2}$ necesariamente sucede que valor-$p > \alpha$, y viceversa. En el caso representado en la figura citada no se rechaza H_0 con cualquiera de los dos criterios. La comparación de t_0 frente a $t_{\alpha/2}$ consiste en contrastar simples números, mientras que comparar las significancias α frente a valor-*p* es contrastar probabilidades, de aquí que esto último sea más informativo.

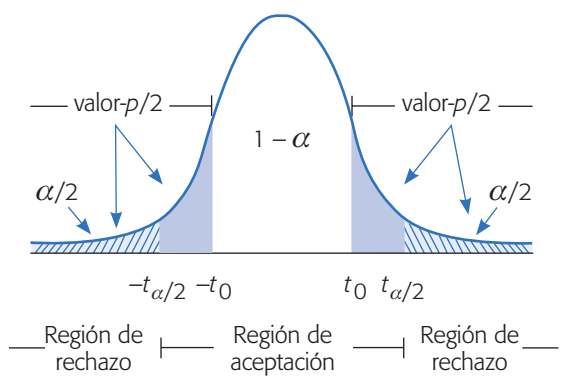


Figura 2.6 Comparación de significancias, valor-$p > \alpha$.

Intervalo de confianza

En este método se rechaza H_0 si el valor del parámetro declarado en la hipótesis nula se encuentra fuera del intervalo de confianza para el mismo parámetro. Cuando la hipótesis planteada es de tipo bilateral, se utiliza directamente el intervalo al 100 $(1-\alpha)\%$ de confianza. Si la hipótesis es unilateral, se requiere el intervalo al 100 $(1-2\alpha)\%$ para que el área bajo la curva, fuera de cada extremo del intervalo, sea igual a α. Por ejemplo, en el caso de la hipótesis unilateral sobre la media del peso de costales dada por la expresión (2.11) se debe construir el intervalo al $100(1-(2\times 0.05))\% = 90\%$ de confianza para aplicar este criterio con una significancia $\alpha = 0.05$. El intervalo al 90% de confianza para la media μ está dado por:

$$\bar{X} \pm t_{0.05}\frac{S}{\sqrt{n}} = 49.40 \pm 1.76\left(\frac{1.095}{3.873}\right) = 49.40 \pm 0.497 = [48.9,\ 49.9]$$

Así, con una confianza de 90% μ está entre 48.9 y 49.9. En tanto, el valor 50.1 declarado en la hipótesis nula no pertenece al intervalo, y además el intervalo está ubicado a la izquierda del 50.1, por lo tanto, se rechaza la hipótesis H_0 : $\mu = 50.1$ y la evidencia señala que contienen menos azúcar de la que se afirma. Nótese que para rechazar una hipótesis unilateral también es necesario verificar la ubicación del intervalo en relación con el valor declarado en la hipótesis nula; el intervalo debe ubicarse con respecto a este valor, como lo indica la hipótesis alternativa. En el ejemplo, la hipótesis alternativa es $H_A : \mu < 50.1$, por lo que para rechazar la hipótesis nula el intervalo debe ubicarse a la izquierda de 50.1, como ocurre en este caso.

Este criterio es útil cuando el software proporciona el intervalo de confianza para el parámetro de interés, pero *no* provee la prueba de hipótesis correspondiente. También puede ser que el experimentador quiera, además de la conclusión de la hipótesis, el intervalo de confianza para el parámetro que le interesa; en ese aspecto, este criterio tiene ventajas sobre los anteriores.

Hipótesis para dos medias: comparación de dos tratamientos

Un problema frecuente que se presenta es comparar la media de dos procesos o dos tratamientos. Por ejemplo, comparar dos proveedores, dos materiales, dos máquinas o dos métodos de trabajo.

Supongamos que interesa comparar dos tratamientos y que éstos son dos máquinas A y B, que realizan la misma operación. Para ello se obtendrá una muestra aleatoria de observaciones de cada máquina. Supongamos que los datos a observar en la máquina A son $Y_{A1}, Y_{A2}, \ldots, Y_{An}$ y los datos de la máquina B son $Y_{B1}, Y_{B2}, \ldots, Y_{Bn}$. Estos futuros datos se podrán escribir como en la tabla 2.2:

Para que la comparación sea justa, la materia prima que utilizan las máquinas se asigna de forma aleatoria a las máquinas, y las $2n$ pruebas o corridas se hacen en orden aleatorio. No es adecuado realizar primero todas las pruebas de la máquina A

Tabla 2.2 Comparación de dos tratamientos.

	Tratamientos	
Prueba o dato	**Máquina A**	**Máquina B**
1	Y_{A1}	Y_{B1}
2	Y_{A2}	Y_{B2}
⋮	⋮	⋮
n	Y_{An}	Y_{Bn}

y posteriormente las de la máquina B, porque eso puede favorecer a una de las máquinas y afecta (sesga) la comparación. La asignación aleatoria del material hace posible que a cada máquina le corresponda material con una calidad equivalente, y el orden aleatorio de las pruebas nulifica el efecto de las fuentes de variabilidad que actúan durante el transcurso de las mismas (como las variables ambientales), al repartir su efecto equitativamente en ambas máquinas. Ahora, veamos cómo hacer estadísticamente este tipo de comparaciones.

Suposición de varianzas desconocidas. Sean dos procesos o tratamientos con medias μ_x y μ_y y varianzas σ_x^2 y σ_y^2, respectivamente. Interesa investigar si las medias de dichos procesos pueden considerarse estadísticamente iguales. Para ello se plantean las siguientes hipótesis:

$$H_0 : \mu_x = \mu_y$$
$$H_A : \mu_x \neq \mu_y \tag{2.12}$$

que se pueden reescribir como

$$H_0 : \mu_x - \mu_y = 0$$
$$H_A : \mu_x - \mu_y \neq 0 \tag{2.13}$$

Para probar H_0 se toman dos muestras aleatorias, como en el ejemplo de las máquinas antes descritas, de tamaño n_x la del proceso X, y de tamaño n_y la del proceso Y; en general, es recomendable que $n_x = n_y = n$, pero también puede trabajarse con $n_x \neq n_y$ si no pudieran tomarse iguales. Si cada proceso sigue una distribución normal y son independientes entre ellos, el estadístico de prueba adecuado para probar la hipótesis de igualdad de medias está dado por,

$$t_0 = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \tag{2.14}$$

el cual sigue una distribución T de Student con $n_x + n_y - 2$ grados de libertad, donde S_p^2 es un estimador de la varianza muestral común, suponiendo que dichas varianzas desconocidas sean iguales, y se calcula como

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

con S_x^2 y S_y^2 las varianzas muestrales de los datos de cada proceso.

Se rechaza H_0 si $|t_0| > t_{\alpha/2}$, donde $t_{\alpha/2}$ es el punto $\alpha/2$ de la cola derecha de la distribución T de Student con $n_x + n_y - 2$ grados de libertad. Cuando la hipótesis alternativa es de la forma $H_A : \mu_x > \mu_y$, se rechaza $H_0 : \mu_x = \mu_y$ si $t_0 > t_\alpha$, y si es de la forma $H_A : \mu_x < \mu_y$, se rechaza si $t_0 < -t_\alpha$. En forma equivalente, se rechaza H_0 si el valor-$p < \alpha$ para la pareja de hipótesis de interés.

Ejemplo 2.5

Comparación de dos centrifugadoras. La calidad de la pintura látex depende, entre otras cosas, del tamaño de la partícula. Para medir esta característica se utilizan dos centrifugadoras, y se sospecha que éstas reportan mediciones distintas para la misma pintura. Se decide hacer un estudio que permita comparar las medias y las varianzas reportadas por los dos equipos; para lo cual, de un mismo lote de pintura se tomaron 13 lecturas con cada centrifugadora. Los resultados son los siguientes:

Centrifugadora *x*	4 714	4 601	4 696	4 896	4 905	4 870	4 987
	5 144	3 962	4 066	4 561	4 626	4 924	
	$\bar{X}_A =$	4 684.00;		$S_A^2 =$	124 732.00		
Centrifugadora *y*	4 295	4 271	4 326	4 530	4 618	4 779	4 752
	4 744	3 764	3 797	4 401	4 339	4 700	
	$\bar{X}_B =$	4 408.92;		$S_B^2 =$	112 020.00		

Para comparar las medias se plantea la hipótesis de igualdad de medias con la alternativa bilateral, puesto que no hay ninguna conjetura del experimentador acerca de cuál centrifugadora puede reportar valores mayores. Luego, el planteamiento es:

$$H_0 : \mu_x = \mu_Y$$
$$H_A : \mu_x \neq \mu_Y$$

la cual se desea probar con un nivel de significancia de 5% ($\alpha = 0.05$). Suponiendo igualdad de varianzas para el tamaño de la partícula, el estadístico de prueba calculado con las fórmulas (2.14) está dado por

$$t_0 = \frac{4\,684.00 - 4\,408.92}{344.06\sqrt{(1/13) + (1/13)}} = 2.04$$

De la tabla de distribución T de Student con $13 + 13 - 2 = 24$ grados de libertad, se obtiene el punto crítico $t_{(0.025,\,24)} = 2.064$. Como $|t_0| = 2.04 < 2.064 = t_{\alpha/2}$, no se rechaza H_0, por lo que se concluye que las centrifugadoras A y B reportan en promedio el mismo tamaño de partícula. Es decir, las centrifugadoras son estadísticamente

iguales en cuanto a sus medias. Sin embargo, conviene observar que el rechazo es por un margen muy escaso, puesto que el estadístico de prueba y el punto crítico son muy similares. Al comparar la significancia predefinida $\alpha = 0.05$ con el valor-p = 0.053 se concluye lo mismo (no se rechaza H_0), pero se aprecia que si tal significancia predefinida por el experimentador fuera $\alpha = 0.055$, la decisión sobre la hipótesis sería rechazada. Esto es, basta con que el experimentador esté dispuesto a correr 0.5% de riesgo tipo I adicional para concluir que las medias de las centrifugadoras no son iguales. En general, no es recomendable cambiar *a posteriori* el valor α para modificar la decisión sobre una hipótesis, pero habría situaciones en las que se pueden admitir probabilidades de este error hasta de $\alpha = 0.1$, dependiendo de lo que implica rechazar la hipótesis.

Otro aspecto a considerar es la significancia práctica de la decisión sobre la hipótesis, lo cual tiene que ver con la diferencia observada, que en este caso es

$$\bar{X} - \bar{Y} = 4\,684.00 - 4\,408.92 = 275.08$$

y representa un estimador de la diferencia en las medias poblacionales del tamaño de partícula que son reportadas por las centrifugadoras. En caso de que 275.08 represente una diferencia relevante, que puede impactar fuertemente la calidad del tamaño de partícula, sería un elemento favorable al tratar de verificar si tal diferencia es real. Ya sea al analizar la conveniencia de utilizar $\alpha = 0.055$ o tomando más datos. Si por el contrario, la diferencia observada se considerara despreciable o irrelevante desde el punto de vista práctico, entonces "conviene" aplicar estrictamente $\alpha = 0.05$ y concluir que las medias de las centrifugadoras son iguales.

El caso de las varianzas desconocidas pero iguales que acabamos de describir, es el más utilizado en la práctica para probar la igualdad de dos medias. En muchos estudios es razonable suponer que las varianzas desconocidas de los dos tratamientos a comparar son iguales. Pero en ocasiones las varianzas no son iguales, o no existen datos históricos sobre los dos tratamientos que permitan suponer algo pertinente sobre las varianzas. Por ejemplo, al comparar dos proveedores del mismo material puede no haber razones para suponer de antemano que las varianzas de cada uno de ellos sean iguales o parecidas (estadísticamente). Si no se supone igualdad de varianzas, el estadístico de prueba para $H_0: \mu_x = \mu_y$ está dado por

$$t_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \tag{2.15}$$

que sigue aproximadamente una distribución T de Student, cuyos grados de libertad v (nu) se calculan mediante la relación:

$$v = \frac{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}\right)^2}{\frac{(S_x^2/n_x)^2}{n_x + 1} + \frac{(S_y^2/n_y)^2}{n_y + 1}} - 2 \tag{2.16}$$

Como antes, se rechaza H_0 si $|t_0| > t_{\alpha/2,\,v}$ o si el valor-$p < \alpha$. Por ejemplo, si en el caso de las centrifugadoras no se pudiera suponer que las varianzas son iguales, el valor del estadístico dado por la expresión (2.15) resulta ser $t_0 = 2.04$, y aplicando la fórmula (2.16) para calcular los grados de libertad, se obtiene que $v = 26$. Con esto se determina que el valor-$p = 0.052$. Por lo tanto, con $\alpha = 0.05$, no se rechaza la igualdad de las medias. Ésta es la misma decisión que se obtuvo al suponer varianzas iguales; observe que los valores de t_0 y el valor-p son prácticamente iguales que antes.

Prueba para la igualdad de varianzas

En lugar de suponer, en la prueba de medias, que las varianzas son iguales o diferentes, se puede proceder a verificarlo de manera estadística mediante las hipótesis:

$$H_0 : \sigma_x^2 = \sigma_y^2$$
$$H_A : \sigma_x^2 \neq \sigma_y^2 \qquad \textbf{(2.17)}$$

La comparación de varianzas tiene interés en sí misma, con independencia de las medias, puesto que éstas son determinantes en cualquier proceso o tratamiento. En general se considera que a menor varianza, implica potencialmente mejor calidad. Por ejemplo, en el caso de las centrifugadoras interesa ver si alguna de ellas tiene mayor error (variabilidad) en sus mediciones. El planteamiento de la hipótesis se puede reescribir como:

$$H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1$$
$$H_A : \frac{\sigma_x^2}{\sigma_y^2} \neq 1 \qquad \textbf{(2.18)}$$

para enfatizar que la prueba se basa en la distribución del estadístico,

$$F_0 = \frac{S_x^2}{S_y^2} \qquad \textbf{(2.19)}$$

el cual, bajo el supuesto de que H_0 es verdad, sigue una distribución F con $n_x - 1$ grados de libertad en el numerador y $n_y - 1$ grados de libertad en el denominador. Al calcular el valor del estadístico de prueba se obtiene que $F_0 = 1.11$. Como la distribución F no es simétrica, el valor-p está dado por el área bajo la curva a la derecha de 1.11, más el área bajo la curva a la izquierda de $1/1.11 = 0.9$.[2] Mediante el paquete estadístico *Statgraphics* se obtiene valor-$p = 0.85$. Por lo tanto, utilizando $\alpha = 0.05$, la decisión es no rechazar $H_0 : \sigma_x^2 = \sigma_y^2$, y se concluye que, estadísticamente,

[2] En general, los puntos porcentuales de cola izquierda y cola derecha de la distribución F cumplen la igualdad: $F_{1-\alpha,\,n_1-1,\,n_2-1} = 1/F_{\alpha,\,n_2-1,\,n_1-1}$. Es decir, uno es el inverso del otro, intercambian los grados de libertad del numerador y del denominador. Si éstos son iguales simplemente es el inverso.

las centrifugadoras tienen la misma variabilidad, precisión o error de medición. El valor del valor-p tan grande con respecto al valor de α, señala que la decisión de no rechazar la igualdad de varianzas es contundente.

Comparación de proporciones

Una situación de frecuente interés es investigar la igualdad de las proporciones de dos poblaciones o tratamientos, es decir, se requiere probar la siguiente hipótesis:

$$H_0 : p_1 = p_2$$
$$H_A : p_1 \neq p_2$$

donde p_1 y p_2 son las proporciones de cada una de las poblaciones o tratamientos. Por ejemplo, para evaluar dos fármacos contra cierta enfermedad se integran dos grupos formados por dos muestras aleatorias de $n_1 = n_2 = 100$ personas cada una. A cada grupo se le suministra un fármaco diferente. Transcurrido el tiempo de prueba se observan $x_1 = 65$ y $x_2 = 75$ personas que se recuperaron con el fármaco en los grupos correspondientes. Para ver si estas diferencias son significativas a favor del fármaco 2, se necesita probar la hipótesis de igualdad de proporciones. Para ello, bajo el supuesto de distribución binomial, el estadístico de prueba z_0 está dado por:

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

donde $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$. Se rechaza H_0 si $|z_0| > z_{\alpha/2}$. En caso de que la hipótesis alterativa fuera unilateral, entonces z_0 se compara con z_α. En el caso de los fármacos, como $\hat{p} = (65 + 75)/(100 + 100) = 0.70$; entonces,

$$z_0 = \frac{\frac{65}{100} - \frac{75}{100}}{\sqrt{0.7(1-0.7)\left(\frac{1}{100} + \frac{1}{100}\right)}} = -1.543$$

Como $|z_0| = 1.543$ no es mayor que $z_{0.0/2} = 1.96$, entonces no se rechaza H_0, por lo que no hay evidencia suficiente para afirmar que un fármaco es mejor que el otro.

Poblaciones pareadas (comparación de dos medias con muestras dependientes)

En las secciones anteriores se probó la hipótesis de igualdad de las medias de dos poblaciones o tratamientos, suponiendo que las dos muestras son independientes. Esta suposición se justifica por la manera en que se obtienen los datos; es decir, a la muestra a la que se le aplica el tratamiento 1 es independiente de la muestra para el tratamiento 2, y los datos se obtienen en orden completamente *al azar*. Con esto se

justifica la suposición de que no existe relación directa entre los datos en el primer tratamiento con los datos en el segundo.

Recordemos que *orden completamente al azar* significa que las unidades se asignan de manera aleatoria a los tratamientos, mientras que las pruebas o corridas experimentales se hacen en orden estrictamente aleatorio, lo cual se hace con la idea de evitar cualquier sesgo que pudiera favorecer a uno de los tratamientos.

Orden completamente al azar
Es aquel en el que las unidades se asignan de manera aleatoria a los tratamientos y las pruebas experimentales se hacen en orden aleatorio.

Sin embargo, en muchas situaciones experimentales no conviene o no es posible tomar muestras independientes, sino que la mejor estrategia es tomar *muestras pareadas*. Esto significa que los datos de ambos tratamientos se van obteniendo por pares, de forma que cada par son datos que tienen algo en común; por ejemplo, que a la misma unidad experimental o espécimen de prueba se le apliquen los tratamientos a comparar. Un par de ejemplos son:

Muestras pareadas
Son aquellas en las que los datos de ambos tratamientos se obtienen por pares, de manera que éstos tienen algo en común y no son independientes.

- A los mismos pacientes se les aplican dos medicamentos (tratamientos) para el dolor en distintas ocasiones; los tratamientos a comparar son los dos medicamentos.
- A las mismas piezas se les hace una prueba de dureza con distintos instrumentos; aquí se quieren comparar los instrumentos.

En el primer caso, el apareamiento consiste en que el grupo de pacientes que recibe el medicamento A es el mismo grupo que recibe el medicamento B, por lo que las mediciones del efecto de los medicamentos sobre el mismo paciente están relacionadas, y en este sentido no son independientes. Al ser el mismo grupo el que recibe ambos tratamientos se logra una comparación más justa y precisa, pero además, al observar las diferencias entre los tratamientos en un mismo paciente se eliminan otras fuentes de variación y se logra hacer una comparación sin sesgos. En el caso de las piezas, si una es grande se espera que ambos instrumentos tiendan a reportar una medición alta, por lo que se espera que haya una fuerte correlación entre las mediciones reportadas con los dos instrumentos. Además, al medir las piezas con los dos instrumentos, si hay diferencias en las mediciones sobre la misma pieza, entonces esas diferencias se deben principalmente al sistema de medición.

Ejemplo 2.6

Comparación de dos básculas. Se desea ver si dos básculas están sincronizadas. Para ello se toma una muestra aleatoria de 10 especímenes y cada uno se pesa en ambas básculas, cuidando que el orden en que se utilizan sea elegido al azar. El trabajo lo realiza el mismo operador y los datos obtenidos se muestran en la tabla 2.3.

Es claro que tenemos el caso de observaciones pareadas, ya que el peso que registra una báscula para un espécimen no es independiente del que registra la otra báscula para el mismo espécimen, en el sentido de que si uno es muy pesado se espera que ambas básculas lo detecten.

La comparación de las básculas se puede evaluar probando la siguiente hipótesis:

$$H_0 : \mu_1 = \mu_2$$
$$H_A : \mu_1 \neq \mu_2$$

Tabla 2.3 Mediciones reportadas por dos básculas.

Espécimen	Báscula 1	Báscula 2	Diferencia
1	11.23	11.27	–0.04
2	14.36	14.41	–0.05
3	8.33	8.35	–0.02
4	10.50	10.52	–0.02
5	23.42	23.41	–0.01
6	9.15	9.17	–0.02
7	13.47	13.52	–0.05
8	6.47	6.46	0.01
9	12.40	12.45	–0.05
10	19.38	19.35	0.03
Medias:	12.87	12.89	–0.022

donde μ_1 es el peso promedio poblacional que mide la báscula 1 y μ_2 es el peso promedio poblacional que mide la báscula 2. Entonces, estas hipótesis, en el caso pareado, se plantean de manera equivalente como:

$$H_0 : \mu_D = 0$$
$$H_A : \mu_D \neq 0 \qquad \textbf{(2.20)}$$

donde μ_D es la media de la población de diferencias. De esta manera, el problema de comparar las medias de dos poblaciones se convierte en el problema de comparar la media de una población con una constante. En este sentido, el estadístico de prueba para la hipótesis (2.20) es el caso particular del estadístico (2.8) para una media, cuando $\mu_0 = 0$ (véase sección "Prueba para la media" de este capítulo). Esto es, con la muestra de n diferencias $(d_1, d_2, \ldots, d_n)$ se obtiene el estadístico dado por:

$$t_0 = \frac{\bar{d}}{S_D/\sqrt{n}} \qquad \textbf{(2.21)}$$

donde $\bar{d} = -0.02$ es el promedio muestral de las diferencias, $S_D = 0.0287$ es la desviación estándar muestral de tales diferencias y n es el tamaño de la muestra. Bajo H_0 el estadístico t_0 se distribuye como una T de Student con $n - 1$ grados de libertad, por lo que H_0 se rechaza si $|t_0| > t_{\alpha/2,\, n-1}$, o si valor-$p < \alpha$. Al hacer los cálculos resulta que:

$$t_0 = \frac{-0.02}{0.0287/\sqrt{10}} = -2.20$$

Como el valor-$p = 0.055$ es mayor que $\alpha = 0.05$ no se rechaza H_0 a un nivel de significancia de $\alpha = 0.05$. Es decir, no hay suficiente evidencia en contra de la sincronización de las básculas. Sin embargo, esta conclusión es bastante endeble dado que el valor-p es muy similar al valor α. De hecho, con $\alpha = 0.06$ se concluiría lo contrario, y el experimentador debería considerar la posibilidad de asumir este riesgo de 6% y rechazar la sincronización de las básculas.

Si en el ejemplo, en lugar de analizar las diferencias, que es la manera correcta de resolver el problema, se analizan por separado las mediciones de cada báscula, el efecto de las posibles diferencias entre los dos instrumentos se mezclaría con la variabilidad que de por sí tienen las piezas. Pero aun si se pesara un grupo diferente de 10 especímenes con cada báscula, entonces la propia variabilidad de las piezas dentro de cada grupo, más las diferencias entre los dos grupos, probablemente ocultaría el efecto de la diferencia de los instrumentos de medición. Así, las observaciones pareadas son necesarias para eliminar fuentes de variabilidad que podrían no dejar hacer la comparación de manera eficaz, esto quedará más claro a continuación.

Poblaciones pareadas: caso más general

La mayoría de las aplicaciones de la prueba pareada buscan una estrategia en donde las diferencias observadas se deban a los tratamientos que se quiere comparar, y no al efecto de la heterogeneidad que de por sí tienen los especímenes de prueba. De esta manera, la prueba pareada puede utilizarse en situaciones más complejas donde es necesario comparar tratamientos ante la presencia de varias fuentes de variabilidad explícitas. Por ejemplo, se quieren comparar dos máquinas por medio de los resultados que generan, pero el material que utilizan tiene una historia larga en la que sufrió el efecto de varios factores como son: proveedores, lotes, turnos, días, subprocesos, etc.; entonces, al no ser posible hacer dos mediciones sobre la misma pieza como en el caso de las básculas, se requiere una identificación más estricta de las fuentes principales de variabilidad a fin de parear los datos con base en ellas.

Ejemplo 2.6

Impurezas en cofres levantados y bajados. En una fábrica de autos se tiene la conjetura o hipótesis de que el número de impurezas en la pintura de los cofres de los autos es diferente, dependiendo de si el auto pasó con el cofre cerrado o abierto por los hornos de secado. Se decide correr un experimento para comparar el número promedio de impurezas en cada situación del cofre (tratamientos). Se consideró que no era adecuado utilizar muestras independientes, ya que se sabía que los días de la semana o los *turnos* podían tener influencia en el número de impurezas. Estos dos factores se incluyen en el estudio como el criterio de apareamiento, como se muestra en la tabla 2.4, en la cual también se aprecian los datos obtenidos. Así, en cada combinación de día y turno se asignaron carros con el cofre levantado y cerrado.

Cada dato en las columnas *levantado* y *bajado* en la tabla 2.4 representa el promedio de impurezas en 10 autos, de tal forma que en el experimento se utilizaron en total 200 autos. La aleatoridad se llevó a cabo por parejas de autos: antes de la entrada a los hornos se aleatorizó si el cofre del primero estaría levantado o bajado; si le tocaba levantado, el cofre del segundo auto debía estar bajado. El planteamiento estadístico consiste en probar la hipótesis de que la media de las diferencias es cero:

$$H_0 : \mu_D = 0$$
$$H_A : \mu_D \neq 0 \qquad \textbf{(2.22)}$$

Tabla 2.4 Número de impurezas en cofres de autos.

Bloque	Día	Turno	Bajado	Levantado	Diferencia
1	LUNES	M	3.4	2.7	0.7
2	LUNES	T	3.7	3.2	0.5
3	MARTES	M	2.9	1.8	1.1
4	MARTES	T	2.5	1.9	0.6
5	MIÉRCOLES	M	1.6	1.1	0.5
6	MIÉRCOLES	T	2.8	2.2	0.6
7	JUEVES	M	3.7	2.8	0.9
8	JUEVES	T	5.9	4.8	1.1
9	VIERNES	M	4.8	4.3	0.5
10	VIERNES	T	4.3	3.4	0.9
		Medias:	3.56	2.82	0.74
		Desviaciones estándar:	1.23	1.15	0.24

Al aceptar $H_0 : \mu_D = 0$ se estaría admitiendo que el número de impurezas promedio en el cofre levantado y bajado son iguales ($H_0 : \mu_L = \mu_B$). El valor del estadístico de prueba es:

$$t_0 = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{0.74}{0.2\,413/\sqrt{10}} = 9.70$$

y el nivel de significancia observado (valor-p) es 0.0000046, el cual es menor que $\alpha = 0.05$, por lo tanto, se rechaza de manera contundente la hipótesis nula de que los tratamientos son iguales, es decir, el número de impurezas en los cofres depende de si éste se encuentra levantado o bajado cuando el auto pasa por los hornos. Pero además, como se observa en los datos, cuando el cofre está levantado hay menos impurezas; entonces, a partir de esto se decidió que los cofres de los autos se levantarán al entrar a los hornos de secado. Con esta medida se logró reducir en forma significativa el número de impurezas.

Nótese que en la tabla 2.4 la gran variabilidad que existe entre los datos de un día a otro, y también entre turnos. Eso causa que, si en lugar de analizar las diferencias se analizan los datos de cada tratamiento (posición del cofre) por separado, las diferencias debido a tratamientos se pueden minimizar ante tanta variabilidad. En efecto, si la comparación se hace siguiendo el criterio de muestras independientes, entonces de acuerdo a lo visto en la sección "Hipótesis para dos medios" de este capítulo, el estadístico de prueba es $t = 1.39$, que le corresponde un valor-$p = 0.18$, por lo que al proceder de esa manera se concluiría en forma equivocada que no hay diferencias entre tratamientos (los detalles de este análisis se dejan como ejercicio). Esto, aunado a las mejoras logradas, justifica que la forma como se hizo el apareamiento fue necesaria y correcta, ya que como se aprecia en el arreglo de la tabla 2.4, se aseguró que al aparear carros pintados el mismo día y en el mismo turno, se logran resultados más homogéneos a los que se les aplican los tratamientos, por lo que las diferencias observadas dentro de un mismo día y turno, se deben en gran medida a los tratamientos.

Resumen de fórmulas para procedimientos de prueba de hipótesis

En la tabla 2.5 se resumen las fórmulas de los procedimientos de pruebas de hipótesis que involucran un parámetro de una sola población, mientras que en la tabla 2.6 se listan los procedimientos que involucran a dos parámetros de dos poblaciones. En cada caso se muestra el planteamiento de la hipótesis, el estadístico de prueba y el criterio de rechazo, este último para cada una de las tres posibles alternativas. Si se trabaja con un software estadístico es más directo y conveniente basarse en el criterio del valor-*p*, el cual, para cualquier hipótesis, debe ser menor que α para que sea posible rechazar H_0.

En la tabla 2.6, note que aparecen tres maneras de probar la igualdad de medias H_0: la primera *a*) es para el caso de muestras independientes suponiendo varianzas iguales. La segunda *b*) es para muestras independientes sin suponer varianzas iguales y el caso *e*) es para muestras pareadas.

Uso de un software estadístico

Los métodos estadísticos tratados en el presente capítulo son más fáciles de aplicar si se utiliza un software para hacer los cálculos. Prácticamente en cualquier software estadístico se incluyen los métodos aquí tratados. Por ejemplo, en *Statgraphics* se incluyen en los menús de *Describe* y *Compare* que aparecen en la pantalla principal. En particular, para hacer una estimación puntual y por intervalo, para la media y la desviación estándar, la secuencia a elegir es la siguiente: *Describe* → *Numeric data* → *One-variable analysis*; entonces, se declara la variable a analizar, la cual fue previamente capturada en una columna de la hoja de datos y después se pide *Confidence intervals* en las opciones tabulares y se especifica el nivel de confianza deseado

Tabla 2.5 Procedimientos de prueba de hipótesis para un parámetro.

Hipótesis	Estadístico de prueba	Criterio de rechazo
a) $H_0 : \mu = \mu_0$	$t_0 = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$	
$H_A : \mu \neq \mu_0$		$\lvert t_0 \rvert > t_{\alpha/2}$
$H_A : \mu > \mu_0$		$t_0 > t_\alpha$
$H_A : \mu < \mu_0$		$t_0 < -t_\alpha$
b) $H_0 : \sigma^2 = \sigma_0^2$	$\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	
$H_A : \sigma^2 \neq \sigma_0^2$		$\chi_0^2 > \chi^2_{\alpha/2,\, n-1}$ o $\chi_0^2 < \chi^2_{1-\alpha/2,\, n-1}$
$H_A : \sigma^2 > \sigma_0^2$		$\chi_0^2 > \chi^2_{\alpha,\, n-1}$
$H_A : \sigma < \sigma_0^2$		$\chi_0^2 < \chi^2_{1-\alpha,\, n-1}$
c) $H_0 : p = p_0$	$z_0 = \dfrac{X - np_0}{\sqrt{np_0(1-p_0)}}$	
$H_A : p \neq p_0$		$\lvert z_0 \rvert > z_{\alpha/2}$
$H_A : p > p_0$	X número de defectos	$z_0 > z_\alpha$
$H_A : p < p_0$		$z_0 < -z_\alpha$

Tabla 2.6 Procedimientos de prueba de hipótesis para dos parámetros.

Hipótesis	Estadístico de prueba	Criterio de rechazo
a) $H_0 : \mu_1 = \mu_2$ $H_A : \mu_1 \neq \mu_2$ $H_A : \mu_1 > \mu_2$ $H_A : \mu_1 < \mu_2$	$t_0 = \dfrac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ donde $S_p = \sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}}$	$\lvert t_0 \rvert > t_{\alpha/2,\, n_1+n_2-2}$ $t_0 > t_{\alpha,\, n_1+n_2-2}$ $t_0 < -t_{\alpha,\, n_1+n_2-2}$
b) $H_0 : \mu_1 = \mu_2$ $H_A : \mu_1 \neq \mu_2$ $H_A : \mu_1 > \mu_2$ $H_A : \mu_1 < \mu_2$	$t_0 = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ donde $v = \dfrac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{(S_1^2/n_1)^2}{n_1+1} + \dfrac{(S_2^2/n_2)^2}{n_2+1}} - 2$	$\lvert t_0 \rvert > t_{\alpha/2,\, v}$ $t_0 > t_{\alpha,\, v}$ $t_0 < -t_{\alpha,\, v}$
c) $H_0 : \sigma_1^2 = \sigma_2^2$ $H_A : \sigma_1^2 \neq \sigma_2^2$ $H_A : \sigma_1^2 > \sigma_2^2$ $H_A : \sigma_1^2 < \sigma_2^2$	$F_0 = \dfrac{S_1^2}{S_2^2}$	$F_0 > F_{\alpha/2,\, n_1-1,\, n_2-1}$ o $F_0 < F_{1-\alpha/2,\, n_1-1,\, n_2-1}$ $F_0 > F_{\alpha,\, n_1-1,\, n_2-1}$ $F_0 < F_{1-\alpha,\, n_1-1,\, n_2-1}$
d) $H_0 : p_1 = p_2$ $H_A : p_1 \neq p_2$ $H_A : p_1 > p_2$ $H_A : p_1 < p_2$	$z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ donde $\hat{p} = \dfrac{X_1 + X_2}{n_1 + n_2}$	$\lvert z_0 \rvert > z_{\alpha/2}$ $z_0 > z_\alpha$ $z_0 < -z_\alpha$
e) $H_0 : \mu_1 = \mu_2$ $H_A : \mu_1 \neq \mu_2$ $H_A : \mu_1 > \mu_2$ $H_A : \mu_1 < \mu_2$	$t_0 = \dfrac{\bar{d}}{S_d / \sqrt{n}}$	$\lvert t_0 \rvert > t_{\alpha/2,\, n-1}$ $t_0 > t_{\alpha,\, n-1}$ $t_0 < -t_{\alpha,\, n-1}$

(*Pane options*). Ahí mismo está la opción *Hypothesis tests*. En las opciones de panel se especifican: el valor (μ_0) que define la hipótesis nula, el nivel de significancia α y el tipo de hipótesis alternativa que se tiene.

Las hipótesis sobre la desviación estándar se prueban en la opción *Confidence intervals* usando el criterio del intervalo de confianza: si el valor especificado en la hipótesis nula σ_0 se encuentra dentro del intervalo no se rechaza H_0; en caso contrario se rechaza.

El problema de comparar dos medias o dos varianzas con muestras independientes, está en *Compare* → *Two samples* → *Two-sample comparison*. En las opciones tabulares se escogen *Comparison of means* y *Comparison of standard deviations*.

Para comparar medias con muestras pareadas la secuencia de opciones a utilizar es: *Compare* → *Two samples* → *Paired-sample comparison*. En Minitab, la secuencia para estimación y prueba de hipótesis es: *Stat* → *Basic Statistics*, y ahí se elige la opción deseada para una, dos muestras (*sample*) o muestra pareada.

En Excel

Para hacer cálculos estadísticos en Excel se utilizan las funciones (f_x) estadísticas y la opción *Análisis de datos* dentro del menú de *Herramientas*. Si no estuviera activada la opción de *Análisis de datos*, ésta se activa usando la opción *Complementos* que está dentro del mismo menú de *Herramientas*. Para probar la hipótesis o encontrar intervalos de confianza para un parámetro, se usa la secuencia: *Herramientas* → *Análisis de datos* → *Estadística descriptiva.* Ahí se activa el cuadro u opción *Nivel de confianza para la media*. En todos los casos, después de señalar el análisis que se desea hacer, se abrirá una ventana en la que se especifica el rango de celdas donde se encuentran los datos y las estadísticas deseadas.

En caso de comparar las medias de dos poblaciones suponiendo varianzas desconocidas pero iguales, la secuencia es: *Herramientas* → *Análisis de datos* → *Prueba t para dos muestras suponiendo varianzas iguales*. Para probar la igualdad de dos medias usando muestras pareadas la secuencia es: *Herramientas* → *Análisis de datos* → *Prueba t para medias de dos muestras emparejadas*. Al final, para probar la igualdad de las varianzas se utiliza la serie de comandos: *Herramientas* → *Análisis de datos* → *Prueba F para varianzas de dos muestras.*

Preguntas y ejercicios

1. En un estudio estadístico, ¿qué es una población y para qué se toma una muestra?
2. ¿Qué es probar una hipótesis?
3. ¿Qué es hacer una estimación puntual y en qué consiste hacer una estimación por intervalo para la media, por ejemplo?
4. ¿Por qué no es suficiente la estimación puntual y por qué se tiene que recurrir a la estimación por intervalo?
5. Explique el papel que desempeñan las distribuciones de probabilidad en la inferencia estadística.
6. En el contexto de estimación por intervalo, señale en forma específica para estimar qué parámetro utiliza cada una de las siguientes distribuciones: *T* de Student, Normal y ji-cuadrada.
7. Explique qué es un estadístico de prueba y señale su relación con los intervalos de aceptación y rechazo.
8. ¿Qué son los errores tipo I y II en pruebas de hipótesis?
9. Señale y describa de manera breve los tres criterios equivalentes de rechazo de una hipótesis.
10. Señale un ejemplo de datos o muestras pareadas.

Ejercicios de estimación

11. En la elaboración de envases de plástico es necesario garantizar que cierto tipo de botella en posición vertical tenga una resistencia mínima de 50 kg de fuerza. Para garantizar esto, en el pasado se realizaba una prueba del tipo pasa-no-pasa, donde se aplicaba la fuerza de 50 kg y se veía si la botella resistía o no. En la actualidad se realiza una prueba exacta, en la que mediante un equipo se aplica fuerza a la botella hasta que ésta cede, y el equipo registra la resistencia que alcanzó la botella.
    a) ¿Qué ventajas y desventajas tiene cada método de prueba?
    b) Para evaluar la resistencia media de los envases se toma una muestra aleatoria de $n = 20$ piezas. De los resultados se obtiene que $\bar{X} = 55.2$ y $S = 3$. Estime con una confianza de 95%, ¿cuál es la resistencia promedio de los envases?
    c) Antes del estudio se suponía que $\mu = 52$. Dada la evidencia de los datos, ¿tal supuesto es correcto?
    d) Con los datos anteriores, estime con una confianza de 95%, ¿cuál es la desviación estándar poblacional (del proceso)?

12. Para evaluar el contenido de nicotina en cierto tipo de cigarros elaborados por un proceso, se toma un muestra aleatoria de 40 cigarrillos y se obtiene que $\bar{X} = 18.1$ mg y $S = 1.7$.
    a) Estime con una confianza de 95%, ¿cuál es la cantidad de nicotina promedio por cigarro?
    b) Antes del estudio se suponía que $\mu = 17.5$. Dada la evidencia de los datos, ¿se puede rechazar tal supuesto?
    c) Con los datos anteriores, estime con una confianza de 95%, ¿cuál es la desviación estándar poblacional (del proceso)?
    d) ¿Qué puede decir sobre la cantidad mínima y máxima de nicotina por cigarro? Es posible garantizar con suficiente confianza que los cigarros tienen menos de 20 mg de nicotina.

13. En un problema similar al del ejercicio 11, es necesario garantizar que la resistencia mínima que tiene un envase de plástico en posición vertical sea de 20 kg. Para evaluar esto se han obtenido los siguientes datos mediante pruebas destructivas:

    28.3 26.8 26.6 26.5 28.1 24.8 27.4 26.2 29.4 28.6 24.9 25.2 30.4 27.7 27.0 26.1 28.1 26.9 28.0 27.6 25.6 29.5 27.6 27.3 26.2 27.7 27.2 25.9 26.5 28.3 26.5 29.1 23.7 29.7 26.8 29.5 28.4 26.3 28.1 28.7 27.0 25.5 26.9 27.2 27.6 25.5 28.3 27.4 28.8 25.0 25.3 27.7 25.2 28.6 27.9 28.7

    a) Esta variable, forzosamente tiene que evaluarse mediante muestreo y no al 100%, ¿por qué?
    b) Haga un análisis exploratorio de estos datos (obtenga un histograma y vea el comportamiento de los datos obtenidos).
    c) Estime, con una confianza de 95%, ¿cuál es la resistencia promedio de los envases?
    d) Antes del estudio se suponía que $\mu = 25$. Dada la evidencia de los datos, ¿tal supuesto es correcto?
    e) Con los datos anteriores estime, con una confianza de 95%, ¿cuál es la desviación estándar poblacional (del proceso)?

14. En la elaboración de una bebida se desea garantizar que el porcentaje de CO (gas) por envase esté entre 2.5 y 3.0. Los siguientes datos son obtenidos del monitoreo del proceso:

2.61 2.62 2.65 2.56 2.68 2.51 2.56 2.62 2.63 2.57 2.60 2.53 2.69 2.53 2.67 2.66 2.63
2.52 2.61 2.60 2.52 2.62 2.67 2.58 2.61 2.64 2.49 2.58 2.61 2.53 2.53 2.57 2.66 2.51
2.57 2.55 2.57 2.56 2.52 2.58 2.64 2.59 2.57 2.58 2.52 2.61 2.55 2.55 2.73 2.51 2.61
2.71 2.64 2.59 2.60 2.64 2.56 2.60 2.57 2.48 2.60 2.61 2.55 2.66 2.69 2.56 2.64 2.67

a) Haga un análisis exploratorio de estos datos (obtenga un histograma y vea el comportamiento de los datos obtenidos).
b) Estime, con una confianza de 95%, ¿cuál es el CO promedio por envase?
c) Se supone que μ debe ser igual a 2.75. Dada la evidencia, ¿se puede rechazar tal supuesto?
d) Con los datos anteriores estime, con una confianza de 95%, ¿cuál es la desviación estándar del proceso?
e) El análisis de los datos muestrales establece que el mínimo es 2.48 y el máximo es 2.73, ¿por qué el intervalo obtenido en el inciso *a*) tiene una menor amplitud?

15. Una característica importante en la calidad de la leche de vaca es la concentración de grasa. En una industria en particular se fijó como estándar mínimo que el producto que recibe directamente de los establos lecheros es de 3.0%. Por medio de 40 muestreos y evaluaciones en cierta época del año se obtuvo que $\bar{X} = 3.2$ y $S = 0.3$.
a) Estime con una confianza de 90% el contenido promedio poblacional de grasa.
b) ¿Cuál es el error máximo de estimación para la media? ¿Por qué?
c) Estime, con una confianza de 95%, ¿cuál es la desviación estándar poblacional?
d) ¿Qué puede decir sobre la cantidad mínima y máxima de grasa en la leche? ¿Es posible garantizar con suficiente confianza que la leche tiene más de 3.0% de grasa?

16. En la fabricación de discos compactos una variable de interés es la densidad mínima (grosor) de la capa de metal, la cual no debe ser menor de 1.5 micras. Se sabe por experiencia que la densidad mínima del metal casi siempre ocurre en los radios 24 y 57, aunque en el método actual también se miden los radios 32, 40 y 48. Se hacen siete lecturas en cada radio dando un total de 35 lecturas, de las cuales sólo se usa la mínima. A continuación se presenta una muestra histórica de 18 densidades mínimas:

1.81, 1.97, 1.93, 1.97, 1.85, 1.99, 1.95, 1.93, 1.85, 1.87, 1.98, 1.93, 1.96, 2.02, 2.07, 1.92, 1.99, 1.93.

a) Argumente estadísticamente si las densidades mínimas individuales cumplen con la especificación de 1.5 micras.
b) Encuentre un intervalo de confianza de 99% para la media de la densidad mínima.
c) Dé un intervalo de confianza de 99% para la desviación estándar.
d) Dibuje el diagrama de cajas para estos datos. Interprete lo que observa.

17. En una auditoría se seleccionan de manera aleatoria 200 facturas de las compras realizadas durante el año, y se encuentra que 10 de ellas tienen algún tipo de anomalía.
a) Estime con una confianza de 95% el porcentaje de facturas con anomalías en todas las compras del año.
b) ¿Cuál es el error de estimación? ¿Por qué?
c) ¿Qué tamaño de muestra se tiene que usar si se quiere estimar tal porcentaje con un error máximo de 2%?

18. En la producción de una planta se está evaluando un tratamiento para hacer que germine cierta semilla. De un total de 60 semillas se observó que 37 de ellas germinaron.
a) Estime con una confianza de 90%, la proporción de germinación que se logrará con tal tratamiento.

b) Con una confianza de 90%, ¿se puede garantizar que la mayoría (más de la mitad) de las semillas germinarán?
c) Conteste los dos incisos anteriores pero ahora con 95% de confianza.

19. Para evaluar la efectividad de un fármaco contra cierta enfermedad se integra en forma aleatoria un grupo de 100 personas. Se suministra el fármaco y transcurrido el tiempo de prueba se observa $x = 65$ personas con un efecto favorable.
a) Estime con una confianza de 90%, la proporción de efectividad que se logrará con tal fármaco. Haga una interpretación de los resultados.

20. En relación al problema del ejercicio 11, los datos anteriores al diseño de la prueba continua muestran lo siguiente: de $n = 120$ envases de plástico probados para ver si tenían la resistencia mínima de 50 kg de fuerza, $x = 10$ envases no pasaron la prueba.
a) Estime con una confianza de 95%, la proporción de envases que no tienen la resistencia mínima especificada. Haga una interpretación de los resultados.
b) ¿Cuál es el error de estimación?
c) Calcule el tamaño de muestra que se necesita para que el error de estimación máximo sea de 0.03.

21. Dos máquinas, cada una operada por una persona, son utilizadas para cortar tiras de hule, cuya longitud ideal debe ser de 200 mm. De las inspecciones de una semana (25 piezas) se observa que la longitud media de las 25 piezas para una máquina es de 200.1 y para la otra es de 201.2. ¿Es significativa la diferencia entre los dos casos? Argumente.

Prueba de hipótesis (comparación de tratamientos)

22. Se desea comprar una gran cantidad de bombillas y se tiene que elegir entre las marcas A y B. Para ello, se compraron 100 focos de cada marca, y se encontró que las bombillas probadas de la marca A tuvieron un tiempo de vida medio de 1 120 horas, con una desviación estándar de 75 horas; mientras que las de la marca B tuvieron un tiempo de vida medio de 1 064 horas, con una desviación estándar de 82 horas.
a) ¿Es significativa la diferencia entre los tiempos medios de vida? Use $\alpha = 0.05$.
b) ¿Con qué tamaño de muestra se aceptaría que las marcas son iguales, utilizando $\alpha = 0.05$?

23. En un laboratorio bajo condiciones controladas, se evaluó, para 10 hombres y 10 mujeres, la temperatura que cada persona encontró más confortable. Los resultados en grados Fahrenheit fueron los siguientes:

Mujer	75	77	78	79	77	73	78	79	78	80
Hombre	74	72	77	76	76	73	75	73	74	75

a) ¿Cuáles son en realidad los tratamientos que se comparan en este estudio?
b) ¿Las muestras son dependientes o independientes? Explique.
c) ¿La temperatura promedio más confortable es igual para hombres que para mujeres? Pruebe la hipótesis adecuada.

24. Se prueban 10 partes diferentes en cada nivel de temperatura y se mide el encogimiento sufrido en unidades de porcentaje multiplicado por 10. Los resultados son:

Temperatura baja	Temperatura alta
17.2	21.4
17.5	20.9
18.6	19.8
15.9	20.4
16.4	20.6
17.3	21.0
16.8	20.8
18.4	19.9
16.7	21.1
17.6	20.3

a) ¿La temperatura tiene algún efecto en el encogimiento? Plantee las hipótesis estadísticas correspondientes a esta interrogante.
b) Dé un intervalo de confianza para la diferencia de medias.
c) ¿Cuál temperatura provoca un encogimiento menor?
d) Compare las varianzas en cada temperatura.
e) Dibuje los diagramas de cajas simultáneos e interprete.

25. Una compañía de transporte de carga desea escoger la mejor ruta para llevar la mercancía de un depósito a otro. La mayor preocupación es el tiempo de viaje. En el estudio se seleccionaron al azar cinco choferes de un grupo de 10 y se asignaron a la ruta A; los cinco restantes se asignaron a la ruta B. Los datos obtenidos fueron:

Ruta	Tiempo de viaje				
A	18	24	30	21	32
B	22	29	34	25	35

a) ¿Existen diferencias significativas entre las rutas? Plantee y pruebe las hipótesis estadísticas correspondientes.
b) En caso de rechazar la hipótesis del inciso *a*), dibuje los diagramas de cajas simultáneos para determinar cuál ruta es mejor.
c) Sugiera otra manera de obtener los datos (diseño alternativo), de manera que se pueda lograr una comparación más efectiva de las rutas.

26. Se tienen dos proveedores de una pieza metálica, cuyo diámetro ideal o *valor objetivo* es igual a 20.25 cm. Se toman dos muestras de 14 piezas a cada proveedor y los datos obtenidos se muestran a continuación:

Proveedor	Diámetros de las piezas de cada proveedor
1	21.38, 20.13, 19.12, 19.85, 20.54, 18.00, 22.24, 21.94, 19.07, 18.60, 21.89, 22.60, 18.10, 19.25
2	21.51, 22.22, 21.49, 21.91, 21.52, 22.06, 21.51, 21.29, 22.71, 22.65, 21.53, 22.22, 21.92, 20.82

a) Describa un procedimiento de aleatorización para la obtención de estos datos.
b) Pruebe la hipótesis de igualdad de los diámetros de los proveedores en cuanto a sus medias.
c) Pruebe la hipótesis de igualdad de varianzas.

d) Si las especificaciones para el diámetro son 20.25 mm ± 2.25 mm, ¿cuál proveedor produce menos piezas defectuosas?
e) ¿Con cuál proveedor se quedaría usted?

27. En Kocaoz, S. Samaranayake, V. A. Nanni A. (2005) se presenta un estudio donde se analizan dos tipos de barras de polímero, cuya tensión se refuerza con fibra de vidrio (FRP). Estas barras, en sustitución de las vigas de acero, son utilizadas para reforzar concreto, por lo que su caracterización es importante para fines de diseño, control y optimización para los ingenieros estructurales. Las barras se sometieron a tensión hasta registrarse su ruptura (en Mpa). Los datos para dos tipos de barras se muestran a continuación:

Tipo de barra	Resistencia							
1	939	976	1 025	1 034	1 015	1 015	1 022	815
2	1 025	938	1 015	983	843	1 053	1 038	938

a) Formule la hipótesis para probar la igualdad de medias de los tratamientos.
b) Anote la fórmula del estadístico de prueba para demostrar la hipótesis.
c) Pruebe la hipótesis a un nivel de significancia de 5%. Para rechazar o no la hipótesis, apóyese tanto en el criterio del valor-*p* como en el del valor crítico de tablas.
d) Explique cómo se obtiene el valor-*p* del inciso anterior.
e) Pruebe la hipótesis de igualdad de varianzas entre tratamientos.
f) ¿Existe algún tratamiento mejor?

28. Se realiza un estudio para comparar dos tratamientos que se aplicarán a frijoles crudos, con el objetivo de reducir el tiempo de cocción. Un tratamiento (T1) es a base de bicarbonato de sodio; el otro, T2, es a base de cloruro de sodio o sal común. La variable de respuesta es el tiempo de cocción en minutos. Se hacen siete réplicas. Los datos se muestran en la siguiente tabla:

Tratamiento	Tiempo						
T1	76	85	74	78	82	75	82
T2	57	67	55	64	61	63	63

a) Formule la hipótesis para probar la igualdad de medias de los tratamientos.
b) Anote la fórmula del estadístico de prueba para probar la hipótesis.
c) Pruebe la hipótesis a un nivel de significancia de 5%. Para rechazar o no la hipótesis, apóyese tanto en el criterio del valor-*p* como en el valor crítico de tablas.
d) Pruebe la hipótesis de igualdad de varianzas entre tratamientos.
e) De acuerdo con el análisis hecho hasta aquí, ¿hay algún tratamiento mejor?

29. Se comparan dos métodos para inocular o contagiar una cepa del hongo del maíz conocido como huitlacoche. En una primera etapa del estudio, el experimentador quiere determinar cuál de los métodos genera mayor porcentaje de infección. El método A consiste en cortar la punta de la mazorca para aplicar la cepa, y en el método B se inyecta la cepa de forma transversal. De 41 mazorcas inoculadas con el método A, 20 se infectaron, es decir, generaron huitlacoche; en tanto, de 38 mazorcas inoculadas con el método B se infectaron 27.
a) ¿Hay evidencia estadística suficiente para afirmar que el método B genera una mayor infección de huitlacoche? Plantee y pruebe la hipótesis correspondiente.

30. El mejor método de inoculación del problema anterior se aplicó a dos variedades de maíz en dos localidades. Una vez infectada la mazorca, interesa medir el porcentaje

final de la superficie de ésta que fue cubierta por el hongo y el peso en gramos del huitlacoche. Los resultados para la variedad 2 de maíz, obtenidos en 15 mazorcas de Texcoco y en 15 mazorcas de Celaya son los siguientes:

Mazorca	% de cobertura (Texcoco)	% de cobertura (Celaya)	Peso en gramos (Texcoco)	Peso en gramos (Celaya)
1	60	95	122.6	231.80
2	40	100	182.74	346.74
3	95	70	203.45	231.41
4	55	40	84.03	141.49
5	40	35	128.46	149.69
6	20	100	31.85	291.28
7	10	30	12.81	86.03
8	10	100	57.05	158.74
9	55	100	145.83	167.25
10	15	100	49.49	120.89
11	35	25	103.66	19.70
12	25	15	95.05	22.08
13	70	85	125.02	134.02
14	20	15	40.57	28.76
15	20	30	19.36	24.87

a) ¿Se puede afirmar que el porcentaje de cobertura del hongo es mayor en Celaya que en Texcoco?
b) Utilice un diagrama de dispersión (gráfica tipo X-Y) para ver si existe una relación lineal entre el porcentaje de cobertura de la mazorca con los gramos de huitlacoche.
c) Ignore la cobertura y pruebe la igualdad de la producción promedio de huitlacoche en las dos localidades.
d) Es evidente que a mayor cobertura hay una mayor producción de huitlacoche, ¿habría forma de saber con estos datos si a igual cobertura corresponde una producción de huitlacoche semejante en ambas localidades? Argumente su respuesta.

31. Con respecto al problema del ejercicio 18, se desea comparar dos tratamientos para hacer que germine cierta semilla. Los datos del tratamiento A son los del ejercicio 18, es decir, de 60 semillas puestas a germinar se observó que 37 de ellas germinaron. Mientras que para el tratamiento B, de 70 semillas se observó que 30 germinaron.
a) ¿Hay una diferencia significativa entre los dos tratamientos? Pruebe la hipótesis correspondiente a 95% de confianza.
b) Estime, con una confianza de 95%, la proporción de germinación que se logrará con cada tratamiento.

32. Se desea comparar dos proveedores; para ello, se toma una muestra aleatoria de la producción de cada uno de $n = 150$ piezas, y se les hace en orden aleatorio una prueba. En el caso del primer proveedor se obtuvieron $x_1 = 11$ piezas que no pasaron la prueba, mientras que para el segundo fueron $x_2 = 22$.
a) ¿Qué proveedor parece mejor?
b) ¿Hay una diferencia significativa entre los dos proveedores? Pruebe la hipótesis correspondiente a 95% de confianza.

Pruebas pareadas

33. La prueba actual de un solo disco se tarda 2 minutos. Se propone un nuevo método de prueba que consiste en medir solamente los radios 24 y 57, donde casi es seguro que estará el valor mínimo buscado. Si el método nuevo resulta igual de efectivo que el método actual se podrá reducir en 60% el tiempo de prueba. Se plantea un experimento donde se mide la densidad mínima de metal en 18 discos usando tanto el método actual como el método nuevo. Los resultados están ordenados horizontalmente por disco. Así 1.88 y 1.87 es el resultado para el primer disco con ambos métodos.

Método actual	1.88	1.84	1.83	1.90	2.19	1.89	2.27	2.03	1.96
	1.98	2.00	1.92	1.83	1.94	1.94	1.95	1.93	2.01
Método nuevo	1.87	1.90	1.85	1.88	2.18	1.87	2.23	1.97	2.00
	1.98	1.99	1.89	1.78	1.92	2.02	2.00	1.95	2.05

a) Pruebe la igualdad de las medias usando la prueba pareada. ¿Cuál es el criterio de apareamiento?
b) Encuentre un intervalo para la diferencia de medias usando la desviación estándar de las diferencias. Inteprete.
c) Haga el análisis de los datos ignorando el apareamiento. Compare con los resultados del inciso *a*), ¿por qué ignorar el apareamiento es incorrecto?
d) Determine un intervalo de confianza para la diferencia de medias suponiendo muestras independientes. Compare con el inciso *b*).
e) ¿Qué se gana con el apareamiento de los datos en este caso?
f) ¿Recomendaría usted la adopción del método nuevo? Argumente su respuesta.

34. En una prueba de dureza, una bola de acero se presiona contra el material al que se mide la dureza. El *diámetro* de la depresión en el material es la medida de su dureza. Se dispone de dos tipos de bolas de acero y se quiere estudiar su desempeño. Para ello, se prueban ambas bolas con los mismos 10 especímenes elegidos de manera aleatoria y los resultados son:

Bola *X*	75	46	57	43	58	32	61	56	34	65
Bola *Y*	52	41	43	47	32	49	52	44	57	60

a) Analice paso a paso cómo se hizo el experimento y explique por qué es importante realizarlo de esa manera.
b) Pruebe la hipótesis de que ambas bolas dan las mismas mediciones de dureza.
c) Pruebe la igualdad de las bolas sin considerar que están pareadas. Compare los resultados con los obtenidos en el inciso *b*).
d) ¿En qué situación se esperaría que los análisis *b*) y *c*) den los mismos resultados?

35. Se conduce un experimento para determinar si el uso de un aditivo químico y un fertilizante estándar aceleran el crecimiento de las plantas. En cada una de 10 localidades se estudiaron dos plantas sembradas en condiciones similares. A una planta de cada localidad se le aplicó el fertilizante puro y a la otra el fertilizante más el aditivo. Después de cuatro semanas el crecimiento en centímetros fue el siguiente:

	Localidad									
	1	2	3	4	5	6	7	8	9	10
Sin aditivo	20	31	16	22	19	32	25	18	20	19
Con aditivo	23	34	15	21	22	31	29	20	24	23

a) ¿Los datos obtenidos apoyan la afirmación de que el aditivo químico acelera el crecimiento de las plantas? Plantee las hipótesis apropiadas y pruébelas usando $\alpha = 0.05$.
b) Obtenga un intervalo al 95% de confianza para la diferencia promedio μ_d.
c) Explique con detalle cómo se pueden asignar de manera aleatoria los tratamientos a las plantas en cada localidad utilizando una moneda.
d) Suponga que en cada localidad una planta queda hacia el Este y la otra hacia el Oeste, realice una asignación aleatoria de los tratamientos a las plantas lanzando una moneda 10 veces.

36. Retome los datos del ejemplo 2.6 (impurezas en cofres levantados y bajados):
a) Ignore el apareamiento, y compare de manera independiente los dos tratamientos. Obtenga conclusiones.
b) Explique si las conclusiones son diferentes con el análisis en forma pareada y de manera independiente.
c) ¿Cuál es la conclusión correcta, hay o no diferencia entre los tratamientos?

37. Se realizó un experimento para ver si dos técnicos tienen alguna tendencia a obtener diferentes resultados cuando determina la pureza de cierto producto. Cada muestra fue dividida en dos porciones y cada técnico determinó la pureza de una de las porciones. Los resultados se muestran a continuación:

	Pureza de las muestras							
Porción	1	2	3	4	5	6	7	8
1	74.0	73.1	73.5	73.9	71.2	72.5	73.0	74.3
2	73.0	71.3	73.2	71.1	70.3	71.5	73.4	72.4

a) Estos datos deben analizarse en forma pareada, explique por qué.
b) Formule la hipótesis correcta al problema.
c) Pruebe la hipótesis y obtenga conclusiones.
d) Si los técnicos son diferentes, ¿hay alguna evidencia sobre cuál de ellos hace mal el trabajo?
e) ¿Qué recomendaría para lograr mayor uniformidad en las determinaciones de los dos técnicos?

Capítulo 3

Experimentos con un solo factor (análisis de varianza)

Sumario

- Diseño completamente al azar y ANOVA
- Comparaciones o pruebas de rango múltiples
- Verificación de los supuestos del modelo
- Elección del tamaño de la muestra
- Uso de software computacional

Objetivos de aprendizaje

- Explicar los elementos de los diseños completamente al azar y el análisis de varianza; asimismo, conocer la importancia del tamaño de la muestra.
- Describir las diversas pruebas de rangos múltiples y la comparación por contrastes.
- Realizar la verificación de los supuestos del modelo.

Mapa conceptual



Conceptos clave

- Análisis de varianza
- Contraste
- Contrastes ortogonales
- Cuadrados medios
- Diagramas de cajas
- Diferencia mínima significativa (LSD)
- Diseño balanceado
- Gráfica de probabilidad en papel normal
- Método de Sheffé
- Métodos de comparaciones múltiples
- Modelo de efectos fijos
- Notación de puntos
- Residuos
- Tabla de análisis de varianza
- Tratamiento control
- Varianza constante

En el capítulo anterior vimos los métodos para comparar dos tratamientos o condiciones (poblaciones o procesos). En este capítulo, aunque se sigue considerando un solo factor, se presentan los diseños experimentales que se utilizan cuando el objetivo es comparar más de dos tratamientos. Puede ser de interés comparar tres o más máquinas, varios proveedores, cuatro procesos, tres materiales, cinco dosis de un fármaco, etcétera.

Es obvio que, al hacer tales comparaciones, existe un interés y un objetivo claro. Por ejemplo, una comparación de cuatro dietas de alimentación en la que se utilizan ratas de laboratorio, se hace con el fin de estudiar si alguna nueva dieta que se propone es mejor o igual que las ya existentes; en este caso, la variable de interés es el peso promedio alcanzado por cada grupo de animales después de ser alimentado con la dieta que le tocó.

Por lo general, el interés del experimentador está centrado en comparar los tratamientos en cuanto a sus medias poblacionales, sin olvidar que también es importante compararlos con respecto a sus varianzas. Así, desde el punto de vista estadístico, la hipótesis fundamental a probar cuando se comparan varios tratamientos es:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu$$
$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j \qquad \textbf{(3.1)}$$

con la cual se quiere decidir si los tratamientos son iguales estadísticamente en cuanto a sus medias, frente a la alternativa de que al menos dos de ellos son diferentes. La estrategia natural para resolver este problema es obtener una muestra representativa de mediciones en cada uno de los tratamientos, y construir un estadístico de prueba para decidir el resultado de dicha comparación.

Se podría pensar que una forma de probar la hipótesis nula de la expresión (3.1) es mediante pruebas T de Student aplicadas a todos los posibles pares de medias; sin embargo, esta manera de proceder incrementaría de manera considerable el error tipo I (rechazar H_0 siendo verdadera). Por ejemplo, supongamos que se desea probar la igualdad de cuatro medias a través de pruebas T de Student. En este caso se tienen seis posibles pares de medias, y si la probabilidad de aceptar la hipótesis nula para cada prueba individual es de $1 - \alpha = 0.95$, entonces la probabilidad de aceptar las seis hipótesis nulas es de $0.95^6 = 0.73$, lo cual representa un aumento considerable del error tipo I. Aunque se utilice un nivel de confianza tal que $(1 - \alpha)^6 = 0.95$, el procedimiento resulta inapropiado porque se pueden producir sesgos por parte del experimentador. Por otra parte, existe un método capaz de probar la hipótesis de igualdad de las k medias con un solo estadístico de prueba, éste es el denominado *análisis de varianza*, el cual se estudiará más adelante.

Diseño completamente al azar y ANOVA

Muchas comparaciones, como las antes mencionadas, se hacen con base en el diseño completamente al azar (DCA), que es el más simple de todos los diseños que se utilizan para comparar dos o más tratamientos, dado que sólo consideran dos fuentes de variabilidad: los *tratamientos* y el *error aleatorio*. En el siguiente capítulo veremos diseños que consideran la influencia de otras fuentes de variabilidad (bloques).

Este diseño se llama *completamente al azar* porque todas las corridas experimentales se realizan en orden aleatorio completo. De esta manera, si durante el estudio se hacen en total N pruebas, éstas se corren al azar, de manera que los posibles efectos ambientales y temporales se vayan repartiendo equitativamente entre los tratamientos.

Ejemplo 3.1

Comparación de cuatro métodos de ensamble. Un equipo de mejora investiga el efecto de cuatro *métodos de ensamble* A, B, C y D, sobre el *tiempo de ensamble* en minutos. En primera instancia, la estrategia experimental es aplicar cuatro veces los cuatro métodos de ensamble en orden completamente aleatorio (las 16 pruebas en orden aleatorio). Los tiempos de ensamble obtenidos se muestran en la tabla 3.1. Si se usa el *diseño completamente al azar* (DCA), se supone que, además del método de ensamble, no existe ningún otro factor que influya de manera significativa sobre la variable de respuesta (tiempo de ensamble).

Más adelante veremos cómo investigar si las diferencias muestrales de la tabla 3.1 garantizan diferencias entre los métodos.

Ejemplo 3.2

Comparación de cuatro tipos de cuero. Un fabricante de calzado desea mejorar la calidad de las suelas, las cuales se pueden hacer con uno de los cuatro tipos de cuero A, B, C y D disponibles en el mercado. Para ello, prueba los cueros con una máquina que hace pasar los zapatos por una superficie abrasiva; la suela de éstos se desgasta al pasarla por dicha superficie. Como criterio de desgaste se usa la pérdida de peso después de un número fijo de ciclos. Se prueban en orden aleatorio 24 zapatos, seis de cada tipo de cuero. Al hacer las pruebas en orden completamente al azar se evitan sesgos y las mediciones en un tipo de cuero resultan independientes de las demás. Los datos (en miligramos) sobre el desgaste de cada tipo de cuero se muestran en la tabla 3.2.

Tabla 3.1 Diseño completamente al azar, ejemplo 3.1.

Método de ensamble			
A	***B***	***C***	***D***
6	7	11	10
8	9	16	12
7	10	11	11
8	8	13	9

Tabla 3.2 Comparación de cuatro tipos de cuero (cuatro tratamientos).

Tipo de cuero	Observaciones						Promedio
A	264	260	258	241	262	255	256.7
B	208	220	216	200	213	206	209.8
C	220	263	219	225	230	228	230.8
D	217	226	215	227	220	222	220.7

Tabla 3.3 Diseño completamente al azar.

Tratamientos				
T_1	T_2	T_3	…	T_k
Y_{11}	Y_{21}	Y_{31}	…	Y_{k1}
Y_{12}	Y_{22}	Y_{32}	…	Y_{k2}
Y_{13}	Y_{23}	Y_{33}	…	Y_{k3}
⋮	⋮	⋮	⋱	⋮
Y_{1n_1}	Y_{2n_2}	Y_{3n_3}	…	Y_{kn_k}

La primera interrogante a despejar es si existen diferencias entre el desgaste promedio de los diferentes tipos de cuero. A continuación veremos la teoría general del diseño y análisis de este tipo de experimentos (DCA), y más adelante se analizarán los datos de los ejemplos planteados.

Supongamos que se tienen k poblaciones o tratamientos, independientes y con medias desconocidas $\mu_1, \mu_2, \ldots, \mu_k$, así como varianzas también desconocidas pero que se suponen iguales $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$. Las poblaciones pueden ser k métodos de producción, k tratamientos, k grupos, etc., y sus medias se refieren o son medidas en términos de la variable de respuesta.

Se decide realizar un experimento completamente al azar para comparar las poblaciones, en principio mediante la hipótesis de igualdad de medias (relación 3.1). Los datos generados por un diseño completamente al azar para comparar dichas poblaciones se pueden escribir como en la tabla 3.3. El elemento Y_{ij} en esta tabla es la j-ésima observación que se hizo en el tratamiento i; n_i es el tamaño de la muestra o las repeticiones observadas en el tratamiento i. Es recomendable utilizar el mismo número de repeticiones ($n_i = n$) en cada tratamiento, a menos que hubiera alguna razón para no hacerlo.[1] Cuando $n_i = n$ para toda i se dice que el *diseño es balanceado*.

Diseño balanceado
Es cuando se utiliza el mismo número de repeticiones en cada tratamiento.

El número de tratamientos k es determinado por el investigador y depende del problema particular de que se trata. El número de observaciones por tratamiento (n) debe escogerse con base en la variabilidad que se espera observar en los datos, así como en la diferencia mínima que el experimentador considera que es importante detectar. Con este tipo de consideraciones, por lo general se recomiendan entre 5 y 30 mediciones en cada tratamiento. Por ejemplo, se usa $n = 10$ cuando las mediciones dentro de cada tratamiento tienen un comportamiento consistente (con poca dispersión). En el otro extremo, se recomienda $n = 30$ cuando las mediciones muestran bastante dispersión. Cuando es costoso o tardado realizar las pruebas para cada tratamiento se puede seleccionar un número menor de repeticiones, con lo cual sólo se podrán detectar diferencias grandes entre los tratamientos.

En caso de que los tratamientos tengan efecto, las observaciones Y_{ij} de la tabla 3.3 se podrán describir con el modelo estadístico lineal dado por:

[1] Si uno de los tratamientos resulta demasiado caro en comparación con los demás, se pueden plantear menos pruebas con éste. Por otra parte, cuando uno de los tratamientos es un control (tratamiento de referencia) muchas veces es el más fácil y económico de probar, y como es de interés comparar a todos los tratamientos restantes con el control, se recomienda realizar más corridas en éste para que sus parámetros queden mejor estimados.

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad \textbf{(3.2)}$$

donde μ es el parámetro de escala común a todos los tratamientos, llamado *media global*, τ_i; es un parámetro que mide el efecto del tratamiento i y ε_{ij} es el error atribuible a la medición Y_{ij}. Este modelo implica que en el diseño completamente al azar actuarían a lo más dos fuentes de variabilidad: los *tratamientos* y el *error aleatorio*. La media global μ de la variable de respuesta no se considera una fuente de variabilidad por ser una constante común a todos los tratamientos, que hace las veces de punto de referencia con respecto al cual se comparan las respuestas medias de los tratamientos (véase figura 3.2). Si la respuesta media de un tratamiento particular μ_i es "muy diferente" de la respuesta media global μ, es un síntoma de que existe un efecto de dicho tratamiento, ya que como se verá más adelante, $\tau_i = \mu_i - \mu$. La diferencia que deben tener las medias entre sí para concluir que hay un efecto (que los tratamientos son diferentes), nos lo dice el análisis de varianza (ANOVA).

En la práctica puede suceder que los tratamientos que se desea comparar sean demasiados como para experimentar con todos. Cuando esto sucede es conveniente comparar sólo una muestra de la población de tratamientos, de modo que τ_i pasa a ser una variable aleatoria con su propia varianza σ_τ^2 que deberá estimarse a partir de los datos (véase sección "Modelos de efectos aleatorios" del capítulo 5). En este capítulo sólo se presenta el caso en que todos los tratamientos que se tienen se prueban, es decir, se supone una población pequeña de tratamientos, lo cual hace posible compararlos a todos. En este caso, el modelo dado por la ecuación (3.2) se llama *modelo de efectos fijos*.

Modelo de efectos fijos
Es cuando se estudian todos los posibles tratamientos.

ANOVA para el diseño completamente al azar (DCA)

El *análisis de varianza* (ANOVA) es la técnica central en el análisis de datos experimentales. La idea general de esta técnica es separar la variación total en las partes con las que contribuye cada fuente de variación en el experimento. En el caso del DCA se separan la variabilidad debida a los tratamientos y la debida al error. Cuando la primera predomina "claramente" sobre la segunda, es cuando se concluye que los tratamientos tienen efecto (figura 3.1*b*), o dicho de otra manera, las medias son diferentes. Cuando los tratamientos no dominan contribuyen igual o menos que el error, por lo que se concluye que las medias son iguales (figura 3.1*a*). Antes de comenzar

Análisis de varianza
Consiste en separar la variación total observada en cada una de las fuentes que contribuye a la misma.

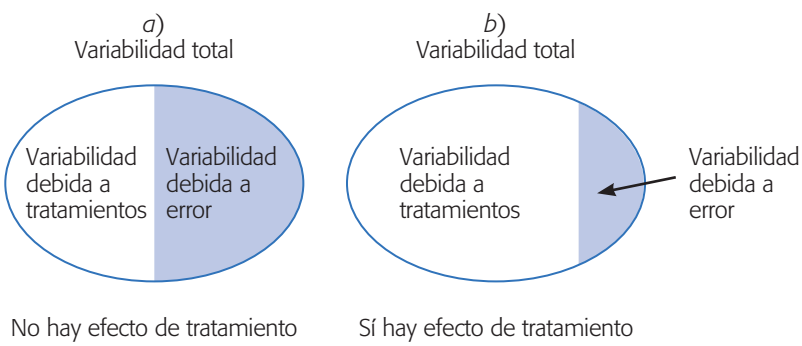


Figura 3.1 Partiendo la variación total en sus componentes en un DCA.

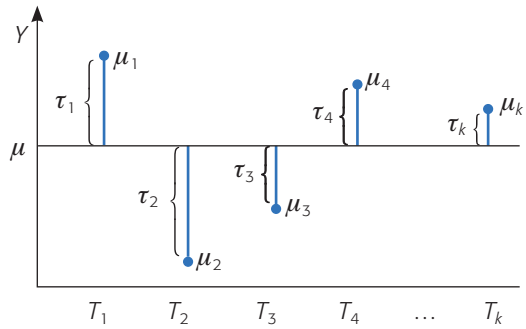


Figura 3.2 Representación de los efectos de los tratamientos en el DCA.

con el análisis del DCA se introduce alguna notación que simplifica la escritura de las expresiones involucradas en dicho análisis.

Notación de puntos
Sirve para representar sumas y medias que se obtienen a partir de los datos experimentales.

Notación de puntos

Sirve para representar de manera abreviada cantidades numéricas que se pueden calcular a partir de los datos experimentales, donde Y_{ij} representa la j-ésima observación en el tratamiento i, con $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, n_i$. Las cantidades de interés son las siguientes:

$Y_{i\bullet}$ = Suma de las observaciones del tratamiento i.
$\bar{Y}_{i\bullet}$ = Media de las observaciones del i-ésimo tratamiento.
$Y_{\bullet\bullet}$ = Suma total de las $N = n_1 + n_2 + \ldots + n_k$ mediciones.
$\bar{Y}_{\bullet\bullet}$ = Media global o promedio de todas las observaciones.

Note que el punto indica la suma sobre el correspondiente subíndice. Así, algunas relaciones válidas son,

$$Y_{i\bullet} = \sum_{j=1}^{n_i} Y_{ij}; \bar{Y}_{i\bullet} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}; Y_{\bullet\bullet} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}$$

$$\bar{Y}_{\bullet\bullet} = \frac{Y_{\bullet\bullet}}{N}; i = 1,2,\ldots,k$$

donde $N = \Sigma_{i=1}^{k} n_i$ es el total de observaciones.

ANOVA

El objetivo del análisis de varianza en el DCA es probar la hipótesis de igualdad de los tratamientos con respecto a la media de la correspondiente variable de respuesta:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu$$
$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j \quad \textbf{(3.3)}$$

la cual se puede escribir en forma equivalente como:

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0$$
$$H_A : \tau_i \neq 0 \text{ para algún } i \quad \textbf{(3.4)}$$

donde τ_i es el efecto del tratamiento i sobre la variable de respuesta. Si se acepta H_0 se confirma que los efectos sobre la respuesta de los k tratamientos son estadísticamente nulos (iguales a cero), y en caso de rechazar se estaría concluyendo que al menos un efecto es diferente de cero.

La equivalencia de las hipótesis (3.3) y (3.4) se deduce directamente del modelo asociado al diseño (ecuación 3.2),[2] pero se observa más fácilmente en la figura 3.2, que es una manera de representar el diseño completamente al azar. En dicha figura se ve que $\tau_i = \mu_i - \mu$, el efecto del tratamiento i, es la distancia entre la respuesta media del tratamiento, μ_i, y la respuesta media global, μ, y cuando un efecto es igual a cero equivale a decir que la media del tratamiento correspondiente es igual a la media global. Así, se observa que para que todas las respuestas medias de tratamientos sean iguales a la respuesta media global μ, representada por la línea horizontal, se requiere que todos los efectos τ_i sean iguales a cero.

Para probar la hipótesis dada por las relaciones (3.3) o (3.4) mediante la técnica de ANOVA, lo primero es descomponer la variabilidad total de los datos en sus dos componentes: la variabilidad debida a tratamientos y la que corresponde al error aleatorio, como se hace a continuación.

Una medida de la variabilidad total presente en las observaciones de la tabla 3.3 es la *suma total de cuadrados* dada por,

$$SC_T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{\bullet\bullet}^2}{N}$$

donde $Y_{\bullet\bullet}$ es la suma de los $N = \Sigma_{i=1}^{n_i} n_i$ datos en el experimento. Al sumar y restar adentro del paréntesis la media del tratamiento i, $(\bar{Y}_{i\bullet})$:

$$SC_T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[(Y_{ij} - \bar{Y}_{i\bullet}) + (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}) \right]^2$$

y desarrollando el cuadrado, la SC_T se puede partir en dos componentes como:

$$SC_T = \sum_{i=1}^{k} n_i (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\bullet})^2$$

donde el primer componente es la *suma de cuadrados de tratamientos* (SC_{TRAT}) y el segundo es la *suma de cuadrados del error* (SC_E). Al observar con detalle estas sumas de cuadrados se aprecia que la SC_{TRAT} mide la variación o diferencias *entre*

[2] Basta observar que $E(Y_{ij}) = \mu + \tau_i = \mu$, de modo que $\tau_i = \mu_i - \mu$.

tratamientos, ya que si éstos son muy diferentes entre sí, entonces la diferencia $\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$ tenderá a ser grande en valor absoluto, y con ello también será grande la SC_{TRAT}. Mientras que la SC_E mide la variación *dentro de tratamientos*, ya que si hay mucha variación entre las observaciones de cada tratamiento entonces $Y_{ij} - \bar{Y}_{i\bullet}$ tenderá a ser grande en valor absoluto. En forma abreviada, esta descomposición de la suma total de cuadrados se puede escribir como:

$$SC_T = SC_{TRAT} + SC_E \tag{3.5}$$

Como hay un total de $N = \Sigma_{i=1}^{n_i} n_i$ observaciones, la SC_T tiene $N - 1$ grados de libertad. Hay k tratamientos o niveles del factor de interés, así que SC_{TRAT} tiene $k - 1$ grados de libertad, mientras que la SC_E tiene $N - k$. Los grados de libertad que corresponden a los términos de la igualdad (3.5) cumplen una relación similar dada por:

$$N - 1 = (k - 1) + (N - k)$$

Cuadrados medios
Es la suma de cuadrados divididos entre sus respectivos grados de libertad.

Las sumas de cuadrados divididas entre sus respectivos grados de libertad se llaman *cuadrados medios*. Los dos que más interesan son el *cuadrado medio de tratamientos* y el *cuadrado medio del error*, que se denotan por

$$CM_{TRAT} = \frac{SC_{TRAT}}{k-1} \text{ y } CM_E = \frac{SC_E}{N-k}$$

Los valores esperados de los cuadrados medios están dados por

$$E(CM_E) = \sigma^2 \text{ y } E(CM_{TRAT}) = \sigma^2 + \frac{\sum_{i=1}^{k} n_i \tau_i^2}{N-k} \tag{3.6}$$

En estas expresiones se aprecia que cuando la hipótesis nula es verdadera, ambos cuadrados medios estiman la varianza σ^2, ya que el segundo término de la expresión para el $E(CM_{TRAT})$ sería igual a cero. Con base en este hecho se construye el estadístico de prueba como sigue: se sabe que SC_E y SC_{TRAT} son independientes, por lo que SC_E/σ^2 y SC_{TRAT}/σ^2 son dos variables aleatorias independientes con distribución ji-cuadrada con $N - k$ y $k - 1$ grados de libertad, respectivamente. Entonces, bajo el supuesto de que la hipótesis H_0 (relaciones 3.3 y 3.4) es verdadera, el estadístico

$$F_0 = \frac{CM_{TRAT}}{CM_E} \tag{3.7}$$

sigue una distribución F con $(k - 1)$ grados de libertad en el numerador y $(N - k)$ grados de libertad en el denominador. De las ecuaciones (3.6) y (3.7) se deduce que

si F_0 es grande, se contradice la hipótesis de que no hay efectos de tratamientos; en cambio, si F_0 es pequeño se confirma la validez de H_0. Así, para un nivel de significancia α prefijado, se rechaza H_0 si $F_0 > F_{\alpha, k-1, N-k}$, donde $F_{\alpha, k-1, N-k}$ es el percentil $(1-\alpha) \times 100$ de la distribución F. También se rechaza H_0 si el valor-$p < \alpha$, donde el valor-p es el área bajo la distribución $F_{k-1, N-k}$ a la derecha del estadístico F_0, es decir, el valor-$p = P(F > F_0)$.

Toda la información necesaria para calcular el estadístico F_0 hasta llegar al valor-p se escribe en la llamada *tabla de análisis de varianza* (ANOVA) que se muestra en la tabla 3.4. En esta tabla, las abreviaturas significan lo siguiente: FV = fuente de variabilidad (efecto), SC = suma de cuadrados, GL = grados de libertad, CM = cuadrado medio, F_0 = estadístico de prueba, valor-p = significancia observada.

Tabla de análisis de varianza
En ésta se resume el análisis de varianza de un experimento, que sirve para probar las hipótesis de interés.

Debemos señalar que el caso particular de comparar dos tratamientos suponiendo varianzas desconocidas pero iguales (prueba T de Student presentada en el capítulo 2), también se puede analizar con el ANOVA y se obtiene el mismo valor del valor-p que con la prueba T. Es fácil comprobar que el estadístico t_0 de la prueba T elevado al cuadrado es igual al estadístico F_0 (3.7) de la prueba F del ANOVA. Por último, es importante resaltar que el ANOVA supone que la variable de respuesta se distribuye *normal*, con *varianza constante* (los tratamientos tienen varianza similar) y que las mediciones son *independientes* entre sí. Estos supuestos deben verificarse para estar más seguros de las conclusiones obtenidas.

Análisis del ejemplo 3.2 (comparación de cuatro tipos de cuero). La interrogante que se planteó en el problema de la comparación entre los cuatro tipos de cuero fue: ¿existen diferencias entre el desgaste promedio de los diferentes tipos de cuero? La respuesta a esta pregunta es el resultado de contrastar las hipótesis:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu$$
$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j \qquad \textbf{(3.8)}$$

En la tabla 3.5 se muestra el análisis de varianza para este ejemplo. Como el valor-p = 0.0000 es menor que la significancia prefijada $\alpha = 0.05$, se rechaza H_0 y se

Tabla 3.4 Tabla de ANOVA para el DCA.

FV	SC	GL	CM	F_0	Valor-p
Tratamientos	$SC_{TRAT} = \sum_{i=1}^{k} \frac{Y_{i\bullet}^2}{n_i} - \frac{Y_{\bullet\bullet}^2}{N}$	$k-1$	$CM_{TRAT} = \frac{SC_{TRAT}}{k-1}$	$\frac{CM_{TRAT}}{CM_E}$	$P(F > F_0)$
Error	$SC_E = SC_T - SC_{TRAT}$	$N-k$	$CM_E = \frac{SC_E}{N-k}$		
Total	$SC_T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{\bullet\bullet}^2}{N}$	$N-1$			

acepta que al menos un par de tipos de cuero tienen un desgaste promedio diferente (la verificación de supuestos se deja al lector como ejercicio).

Si al menos un tipo de cuero se desgasta de forma diferente de otro, entonces ¿cuáles tipos de cuero son diferentes entre sí? Para responder esta pregunta se realizan todas las comparaciones posibles, dos a dos entre las medias de tratamientos, para lo cual existen varios métodos de prueba conocidos genéricamente como *métodos de comparaciones múltiples*, algunos de los cuales se presentan más adelante, junto con otros análisis gráficos que permiten entender mejor los resultados.

Método de comparaciones múltiples
Técnicas para comparar todos los posibles pares de medias de tratamientos.

Además de la tabla 3.5 del ANOVA se observa que la variación total en 24 datos de este experimento fue de 9 101. De esta cantidad, 7 072 se debe a las diferencias entre los tipos de cuero y 2 029 corresponde a la diferencia entre los cueros del mismo tipo. Al ponderar esto por los correspondientes grados de libertad, se obtienen los cuadrados medios que reflejan la magnitud real de cada fuente de variación. Así, vemos que las diferencias debido al tipo de cuero es de 2 357 y que el error es de 101; por lo tanto, la primera es 23.2 veces más grande que la segunda, lo cual indica que las diferencias observadas entre los tipos de cuero son significativas y que no se deben a pequeñas variaciones muestrales (error).

Ejemplo 3.3

Comparación de cuatro métodos de ensamble. Consideremos los datos del DCA dados en el ejemplo 3.1, donde el interés era comparar cuatro métodos de ensamble en cuanto al tiempo promedio en minutos que requiere cada uno de ellos. Se hicieron cuatro observaciones del tiempo de ensamble en cada método. Los resultados se muestran en la tabla 3.1.

Una manera de comparar los métodos de ensamble (tratamientos) es probar la hipótesis:

$$H_0 : \tau_A = \tau_B = \tau_C = \tau_D = 0$$
$$H_A : \tau_i \neq 0 \text{ para algún } i = A, B, C, D \qquad \textbf{(3.9)}$$

En caso de no rechazar H_0 se concluye que los tiempos promedio de los cuatro métodos de ensamble son estadísticamente iguales; pero si se rechaza, se concluye que al menos dos de ellos son diferentes. En la tabla 3.6 se muestra el análisis de varianza correspondiente, en donde se aprecia que el valor del valor-p = 0.0018 es menor que $\alpha = 0.05$, por lo que se rechaza H_0 en este nivel de significancia en particular. No obstante, también se rechazaría para cualquier otro nivel de significancia

Tabla 3.5 ANOVA para los tipos de cuero.

FV	SC	GL	CM	F_0	Valor-p
Tipo de cuero	7 072.33	3	2 357.44	23.24	0.0000
Error	2 029.0	20	101.45		
Total	9 101.33	23			

Tabla 3.6 ANOVA para los métodos de ensamble.

FV	SC	GL	CM	F_0	Valor-*p*
Tratamientos	69.5	3	23.17	9.42	0.0018
Error	29.5	12	2.46		
Total	99.0	15			

prefijado, α, que cumpla con $\alpha > 0.0018$, ya que en esos casos el estadístico de prueba $F_0 = 9.42$ caería en la región de rechazo.

Cálculos manuales

Hay personas que, cuando hacen los cálculos de forma manual, complementan el entendimiento de un análisis con el apoyo de una calculadora de bolsillo, al menos para los casos más simples. Para el caso del ANOVA, estos cálculos se facilitan si primero se obtiene la información básica desplegada en la tabla 3.7. Con esta información se pueden calcular las sumas de cuadrados, como se hace a continuación:

1. Suma total de cuadrados o variabilidad total de los datos:

$$SC_T = \sum_{i=j}^{4}\sum_{j=1}^{4} Y_{ij}^2 - \frac{Y_{\cdot\cdot}^2}{N} = 1\,620 - \frac{156^2}{16} = 99.0$$

2. Suma de cuadrados de tratamientos o variabilidad debida a la diferencia entre métodos de ensamble:

$$SC_{TRAT} = \sum_{i=1}^{4} \frac{Y_{i\cdot}^2}{4} - \frac{Y_{\cdot\cdot}^2}{N} = \frac{(29^2 + 34^2 + 51^2 + 42^2)}{4} - \frac{156^2}{16} = 69.5$$

3. Suma de cuadrados del error o variabilidad dentro de métodos de ensamble:

$$SC_E = SC_T - SC_{TRAT} = 99 - 69.5 = 29.5$$

Tabla 3.7 Detalles de los cálculos para el ANOVA en el DCA para el tiempo de ensamble, ejemplo 3.3.

Métodos de ensamble					Operaciones básicas
	A	*B*	*C*	*D*	
Observaciones ⇒	6	7	11	10	$\Sigma_{i=1}^{4}\Sigma_{j=1}^{4} Y_{ij}^2 = 6^2 + 7^2 + \ldots + 9^2 = 1\,620$
	8	9	16	12	= suma de los cuadrados de todas las observaciones o datos
	7	10	11	11	
	8	8	13	9	
Total por tratamiento $(Y_{i\cdot})$ ⇒	29	34	51	42	$Y_{\cdot\cdot} = \Sigma_{i=1}^{4}\Sigma_{j=1}^{4} Y_{ij} = 6 + 7 + \ldots + 9 = 156$ suma de los datos
Número de datos en cada tratamiento (n_i) ⇒	4	4	4	4	$N = \Sigma_{i=1}^{4} n_i = 16$ total de mediciones
Media muestral por tratamiento $(\bar{Y}_{i\cdot})$ ⇒	7.25	8.50	12.75	10.50	$\bar{Y}_{\cdot\cdot} = \frac{Y_{\cdot\cdot}}{N} = \frac{156}{16} = 9.75$ media global
Desviaciones respecto a la media global $(\hat{\tau}_i)$ ⇒	–2.50	–1.25	3.0	0.75	$\hat{\tau}_i = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$ efecto estimado del método *i*

4. Cuadrados medios de tratamientos y del error (efecto ponderado de cada fuente de variación):

$$CM_{TRAT} = \frac{SC_{TRAT}}{k-1} = \frac{69.5}{3} = 23.17 \text{ y } CM_E = \frac{SC_E}{N-k} = \frac{29.5}{12} = 2.46$$

5. Estadístico de prueba:

$$F_0 = \frac{CM_{TRAT}}{CM_E} = \frac{23.17}{2.46} = 9.42$$

Con toda esta información se procede a llenar la tabla 3.6 de ANOVA. El valor de la significancia observada o valor-*p* es el área bajo la curva de la distribución $F_{3,\,12}$ a la derecha de $F_0 = 9.42$, lo cual es difícil de calcular de forma manual. Sin embargo, cuando esto no sea posible, recordemos que otra forma de rechazar o no una hipótesis es comparar el estadístico de prueba contra un número crítico de tablas. En el caso de las tablas de la distribución *F* en el apéndice, se lee que el valor crítico para $\alpha = 0.05$ es $F_{0.05,\,3,\,12} = 3.49$. Como $F_0 = 9.42 > F_{0.05,\,3,\,12} = 3.49$, entonces se rechaza H_0, con lo cual se concluye que sí hay diferencia o efecto de los métodos de ensamble en cuanto a su tiempo promedio.

Diagramas de cajas simultáneos

Diagramas de caja
Gráficos basados en los cuartiles de un conjunto de datos.

Los *diagramas de cajas*[3] *simultáneos* representan una manera descriptiva de comparar tratamientos. En la figura 3.3 se presentan los diagramas de cajas simultáneos para los cuatro métodos de ensamble del ejemplo 3.3. Se observa que el método *C* parece diferente a los métodos *A* y *B* en cuanto a sus medias; la media del método *D* también se ve diferente a la media del método *A*. Por otra parte, se observa un poco más de variabilidad en el método *C* que en todos los demás. Lo que sigue es verificar que lo que se observa en el diagrama de caja implica diferencias significativas entre los distintos tratamientos; por lo tanto, es necesario hacer pruebas estadísticas porque los datos que se analizan en los diagramas de cajas son muestras.

En general, cuando los diagramas no se traslapan es probable que los tratamientos correspondientes sean diferentes entre sí, y la probabilidad es mayor en la medida que los diagramas están basados en más datos. Cuando se traslapan un poco puede ser que haya o no diferencias significativas, y en cualquier caso es conveniente utilizar una prueba estadística para determinar cuáles diferencias son significativas. Estas pruebas se verán en la siguiente sección.

Gráficos de medias

Cuando se rechaza H_0 mediante el ANOVA, y se concluye que no hay igualdad entre las medias poblacionales de los tratamientos, pero no se tiene información específica

[3] El diagrama de caja es una herramienta para describir el comportamiento de unos datos, y es de suma utilidad para comparar procesos, tratamientos y, en general, para hacer análisis por estratos (lotes, proveedores, turnos). El diagrama de caja se basa en los cuartiles y parte el rango de variación de los datos en cuatro grupos, cada uno de los cuales contiene 25% de las mediciones. De esta forma se puede visualizar dónde empieza 25% de los datos mayores, dónde 25% de los datos menores y de dónde a dónde se ubica 50% de los datos que están al centro.

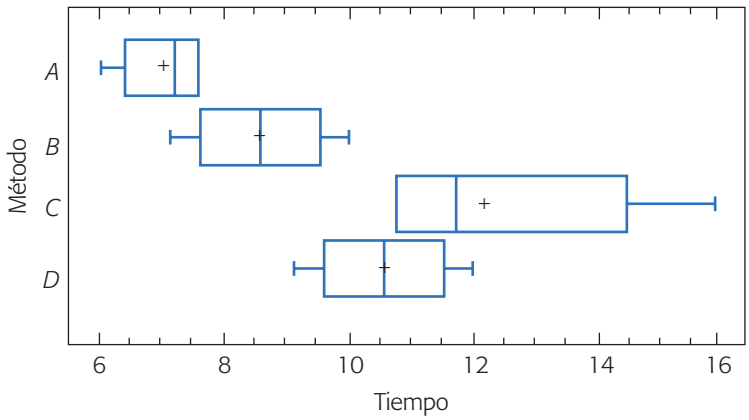


Figura 3.3 Diagramas de cajas para los métodos de ensamble.

sobre cuáles tratamientos son diferentes entre sí, el gráfico de medias (*means plot*) permite hacer una comparación visual y estadística de las medias de los tratamientos (métodos de ensamble). En la figura 3.4 se presenta el gráfico de medias con intervalos de confianza de acuerdo con la prueba LSD, la cual se estudiará más adelante.

Como se explicó en el capítulo anterior, si dos intervalos de confianza se traslapan, los tratamientos correspondientes son estadísticamente iguales en cuanto a sus medias; pero si no se traslapan, entonces son diferentes. Así, podemos ver que el método LSD detecta con una confianza de 95% que $A \neq C$, $A \neq D$ y $B = C$. De esta forma, la conclusión práctica del experimento es que el mejor método de ensamble parece ser el *A*, ya que estadísticamente sus tiempos son menores que los de los métodos *C* y *D*. Le sigue el método *B*, ya que éste es mejor que el *C*. Pero no es posible concluir que el método *A* sea mejor que el método *B*, ya que sus intervalos se traslapan. Si se quisiera decidir en forma estadística sobre la diferencia entre los métodos *A* y *B*, una forma de hacerlo es tomar más datos para incrementar la potencia de la prueba, o bien, recurrir a otros criterios para tomar la decisión.

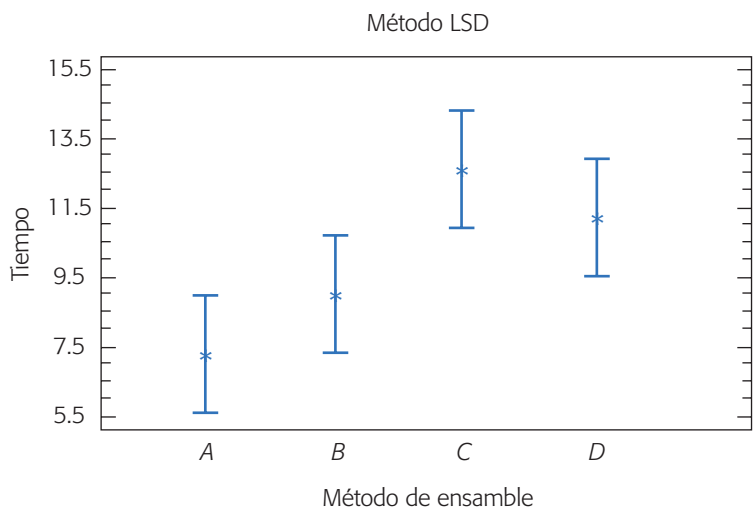


Figura 3.4 Gráfico de medias con el método LSD (ejemplo 3.3).

Comparaciones o pruebas de rango múltiples

Después de que se rechazó la hipótesis nula en un análisis de varianza, es necesario ir a detalle y ver cuáles tratamientos son diferentes. A continuación veremos tres estrategias distintas para ir a ese detalle.

Comparación de parejas de medias de tratamientos

Cuando no se rechaza la hipótesis nula $H_0 : \mu_1 = \mu_2 = \ldots \mu_k = \mu$, el objetivo del experimento está cubierto y la conclusión es que los tratamientos no son diferentes. Si por el contrario se rechaza H_0, y por consiguiente se acepta la hipótesis alternativa $H_A : \mu_i \neq \mu_j$ para algún $i \neq j$, es necesario investigar cuáles tratamientos resultaron diferentes, o cuáles provocan la diferencia. Como se acaba de ilustrar en la gráfica de medias, estas interrogantes se responden probando la igualdad de todos los posibles pares de medias, para lo que se han propuesto varios métodos, conocidos como *métodos de comparaciones múltiples* o *pruebas de rango múltiple*. La diferencia primordial entre los métodos radica en la potencia que tienen para detectar las diferencias entre las medias. Se dice que una prueba es más potente si es capaz de detectar diferencias más pequeñas.

Diferencia mínima significativa (LSD)
Es la diferencia mínima que debe haber entre dos medias muestrales para considerar que dos tratamientos son diferentes.

Método LSD (diferencia mínima significativa)

Una vez que se rechazó H_0 en el ANOVA, el problema es probar la igualdad de todos los posibles pares de medias con la hipótesis:

$$\begin{aligned} H_0 &: \mu_i = \mu_j \\ H_A &: \mu_i \neq \mu_j \end{aligned} \tag{3.10}$$

para toda $i \neq j$. Para k tratamientos se tienen en total $k(k-1)/2$ pares de medias. Por ejemplo, si $k = 4$ existen $4 \times 3/2 = 6$ posibles pares de medias. El estadístico de prueba para cada una de las hipótesis dadas en (3.11) es la correspondiente diferencia en valor absoluto entre sus medias muestrales $\left|\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\right|$. Se rechaza la hipótesis $H_0 : \mu_i = \mu_j$ si ocurre que

$$\left|\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\right| > t_{\alpha/2,\,N-k}\sqrt{CM_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} = LSD \tag{3.11}$$

donde el valor de $t_{\alpha/2,\,N-k}$ se lee en las tablas de la distribución T de Student con $N-k$ grados de libertad que corresponden al error, el CM_E es el cuadrado medio del error y se obtiene de la tabla de ANOVA, n_i y n_j son el número de observaciones para los tratamientos i y j, respectivamente. La cantidad LSD se llama *diferencia mínima*

significativa (*least significant difference*), ya que es la diferencia mínima que debe existir entre dos medias muestrales para considerar que los tratamientos correspondientes son significativamente diferentes. Así, cada diferencia de medias muestrales en valor absoluto que sea mayor que el número LSD se declara significativa. Note que si el diseño es balanceado, es decir, si $n_1 = n_2 = \ldots = n_k = n$, la diferencia mínima significativa se reduce a:

$$LSD = t_{\alpha/2,\ N-k}\sqrt{2CM_E/n} \tag{3.12}$$

En caso de rechazar H_0 se acepta la hipótesis alternativa $H_A : \mu_i \neq \mu_j$, la cual nos dice que las medias de los tratamientos i y j son diferentes. El método LSD tiene una potencia importante, por lo que en ocasiones declara significativas aun pequeñas diferencias.

Ejemplo 3.4

Ilustremos esta prueba continuando con el ejemplo 3.3, en el cual, con el ANOVA se rechazó la hipótesis $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$ y se acepta que al menos un par de medias de tratamientos (métodos de ensamble) son diferentes entre sí. Para investigar cuáles pares de medias son estadísticamente diferentes se prueban los seis posibles pares de hipótesis:

$$\begin{aligned}
&H_0 : \mu_A = \mu_B \quad vs. \quad H_A : \mu_A \neq \mu_B \\
&H_0 : \mu_A = \mu_C \quad vs. \quad H_A : \mu_A \neq \mu_C \\
&H_0 : \mu_A = \mu_D \quad vs. \quad H_A : \mu_A \neq \mu_D \\
&H_0 : \mu_B = \mu_C \quad vs. \quad H_A : \mu_B \neq \mu_C \\
&H_0 : \mu_B = \mu_D \quad vs. \quad H_A : \mu_B \neq \mu_D \\
&H_0 : \mu_C = \mu_D \quad vs. \quad H_A : \mu_C \neq \mu_D
\end{aligned} \tag{3.13}$$

utilizando el método de LSD. En el ANOVA de la tabla 3.6 se observa que los grados de libertad del error son $N - k = 12$, y que el cuadrado medio del error es $CM_E = 2.46$. Si usamos una significancia predefinida de $\alpha = 0.05$, de la tabla de la distribución T de Student con 12 grados de libertad, se obtiene que $t_{0.025,\ 12} = 2.18$. Como en cada tratamiento se hicieron $n = 4$ pruebas, entonces:

$$LSD = t_{\alpha/2,\ N-k}\sqrt{2CM_E/n} = 2.18\sqrt{\frac{2 \times 2.46}{4}} = 2.42$$

La decisión sobre cada una de las seis hipótesis listadas arriba se obtiene al comparar las correspondientes diferencias de medias muestrales en valor absoluto con el número LSD = 2.42. Se declaran significativas aquellas diferencias que son mayores a este número. Los resultados se muestran en la tabla 3.8, de donde se concluye que $\mu_A = \mu_B, \mu_B = \mu_D, \mu_C = \mu_D$, mientras que $\mu_A \neq \mu_C, \mu_B \neq \mu_C$ y $\mu_A \neq \mu_D$. Note que son los mismos resultados que previamente se obtuvieron en la gráfica de medias (figura 3.4), cuyos intervalos están basados en este método LSD. De manera

Tabla 3.8 Aplicación de la prueba *LSD* a métodos de ensamble.

Diferencia poblacional	Diferencia muestral en valor absoluto	Decisión
$\mu_A - \mu_B$	$1.25 < 2.42$	No significativa
$\mu_A - \mu_C$	$* 5.50 > 2.42$	Significativa
$\mu_A - \mu_D$	$* 3.25 > 2.42$	Significativa
$\mu_B - \mu_C$	$* 4.25 > 2.42$	Significativa
$\mu_B - \mu_D$	$2.00 < 2.42$	No significativa
$\mu_C - \mu_D$	$2.25 < 2.42$	No significativa

específica, los intervalos en la gráfica de medias (*means plot*) con el método LSD se obtienen con:

$$\bar{Y}_{i\bullet} \pm t_{\alpha/2,\, N-k}\sqrt{\frac{CM_E}{n_i}}$$

De esta forma, si dos intervalos se traslapan, entonces no habrá diferencias entre las medias de los tratamientos correspondientes. Note que $\sqrt{CM_E/n}$ se está considerando como el error estándar o desviación estándar de la correspondiente media muestral.

Método de Tukey

Un método más conservador para comparar pares de medias de tratamientos es el *método de Tukey*, el cual consiste en comparar las diferencias entre medias muestrales con el valor crítico dado por:

$$T_\alpha = q_\alpha\,(k, N-k)\ \sqrt{CM_E/n_i}$$

donde CM_E es el cuadrado medio del error, n es el número de observaciones por tratamiento, k es el número de tratamientos, $N - k$ es igual a los grados de libertad para el error, α es el nivel de significancia prefijado y el estadístico $q_\alpha(k, N - k)$ son puntos porcentuales de la distribución del rango estudentizado, que se obtienen de la correspondiente tabla en el apéndice. Se declaran significativamente diferentes los pares de medias cuya diferencia muestral en valor absoluto sea mayor que T_α. A diferencia de los métodos LSD y Duncan, el método de Tukey trabaja con un error α muy cercano al declarado por el experimentador.

Ejemplo 3.5

Para aplicar el método de Tukey al ejemplo de los métodos de ensamble, a partir del ANOVA de la tabla 3.6, se toma la información pertinente y de las tablas del rango estudentizado dadas en el apéndice, para $\alpha = 0.05$, se obtiene $q_{0.05}(4, 12) = 4.20$, de manera que el valor crítico es:

$$T_{0.05} = q_{0.05}(4, 12)\sqrt{CM_E/n} = 4.20 \times \sqrt{2.46/4} = 3.27$$

que al compararlo con las diferencias de medias muestrales, los resultados sobre las seis hipótesis son:

Diferencia poblacional	Diferencia muestral	Decisión
$\mu_A - \mu_B$	$1.25 < 3.27$	No significativa
$\mu_A - \mu_C$	$* 5.50 > 3.27$	Significativa
$\mu_A - \mu_D$	$3.25 > 3.27$	No significativa
$\mu_B - \mu_C$	$* 4.25 > 3.27$	Significativa
$\mu_B - \mu_D$	$2.00 < 3.27$	No significativa
$\mu_C - \mu_D$	$2.25 < 3.27$	No significativa

De esta tabla se concluye que $\mu_A = \mu_B = \mu_D$, $\mu_C = \mu_D$, $\mu_A \neq \mu_C$ y $\mu_B \neq \mu_C$. Observe que esta prueba no encuentra diferencia entre los métodos de ensamble *A* y *D*, la cual sí se detectó con el método LSD. Esto es congruente con el hecho de que la prueba de Tukey es menos potente que la prueba LSD, por lo que las pequeñas diferencias no son detectadas como significativas. Asimismo, el riesgo de detectar una diferencia que no existe es menor con el método de Tukey. En la práctica, después de que se ha rechazado H_0 con el ANOVA, conviene aplicar ambos métodos (LSD y Tukey) u otros, cuando haya dudas sobre cuál es el tratamiento ganador. Cuando la diferencia entre dos tratamientos es clara, ambos métodos coinciden.

Método de Duncan

En este método para la comparación de medias, si las k muestras son de igual tamaño, los k promedios se acomodan en orden ascendente y el error estándar de los promedios se estima con $S_{\bar{Y}_{i\cdot}} = \sqrt{CM_E/n}$. Si alguno o todos los tratamientos tienen tamaños diferentes, se reemplaza n por la media armónica de las $\{n_i\}$, que está dada por,

$$n_{AR} = \frac{k}{\sum_{i=1}^{k} \frac{1}{n_i}}$$

Nótese que cuando $n_1 = n_2 = \ldots = n_k = n$, ocurre que $n_{AR} = n$. De la tabla de rangos significantes de Duncan dada en el apéndice, se obtienen los valores críticos $r_\alpha(p, l)$, $p = 2, 3, \ldots, k$, donde α es el nivel de significancia prefijado y l son los grados de libertad para el error. Con estos $k - 1$ valores se obtienen los rangos de significancia mínima dados por

$$R_p = r_\alpha(p, l) S_{\bar{Y}_{i\cdot}};\ p = 2, 3, \ldots, k$$

Las diferencias observadas entre las medias muestrales se comparan con los rangos R_p de la siguiente manera: primero se compara la diferencia entre la media más grande y la más pequeña con el rango R_k. Luego, la diferencia entre la media más grande y la segunda más pequeña se compara con el rango R_{k-1}. Estas comparaciones continúan hasta que la media mayor se haya comparado con todas las demás. Enseguida, se compara la diferencia entre la segunda media más grande y la media

menor con el rango R_{k-1}. Después, la diferencia entre la segunda media más grande y la segunda más pequeña se compara con el valor de R_{k-2}, y así sucesivamente hasta que se comparan los $k(k-1)/2$ pares de medias posibles con el rango que les corresponda. En las comparaciones donde la diferencia observada es mayor que el rango respectivo, se concluye que esas medias son significativamente diferentes. Si dos medias caen entre otras dos que no son muy diferentes, entonces esas dos medias poblacionales también se consideran estadísticamente iguales.

Ejemplo 3.6

De nuevo, supongamos que interesa probar las seis hipótesis dadas en (3.13) para los cuatro métodos de ensamble. En la tabla de ANOVA (tabla 3.6) se lee que $CM_E = 2.46$, lo cual se basa en 12 grados de libertad. Así, el error estándar de cada promedio es $S_{\bar{Y}_{i\bullet}} = \sqrt{CM_E/n} = \sqrt{2.46/4} = 0.78$, dado que se hicieron $n = 4$ observaciones en cada tratamiento. De la tabla de rangos significantes de Duncan dada en el apéndice, para $\alpha = 0.05$ y 12 grados de libertad, se leen los rangos $r_{0.05}(2, 12) = 3.08$, $r_{0.05}(3, 12) = 3.23$ y $r_{0.05}(4, 12) = 3.33$. Con esta información, los rangos mínimos significantes son:

$$R_2 = r_{0.05}(2, 12)S_{\bar{Y}_{i\bullet}} = (3.08)(0.78) = 2.40$$
$$R_3 = r_{0.05}(3, 12)S_{\bar{Y}_{i\bullet}} = (3.23)(0.78) = 2.52$$
$$R_4 = r_{0.05}(4, 12)S_{\bar{Y}_{i\bullet}} = (3.33)(0.78) = 2.60$$

Estos rangos se comparan con las diferencias de medias de acuerdo al método descrito arriba.

Las cuatro medias muestrales acomodadas en orden ascendente son: $\bar{Y}_A = 7.25$, $\bar{Y}_B = 8.50$, $\bar{Y}_D = 10.50$ y $\bar{Y}_C = 12.75$. De aquí se obtienen las diferencias en el orden dado por el método de Duncan y se van comparando con el rango correspondiente. En la siguiente tabla se resumen los resultados obtenidos.

Diferencia poblacional	Diferencia muestral comparada con su rango R_p	Decisión
$\mu_C - \mu_A$	$12.75 - 7.25 = 5.5^* > 2.60 = R_4$	Significativa
$\mu_C - \mu_B$	$12.75 - 8.50 = 3.27^* > 2.52 = R_3$	Significativa
$\mu_C - \mu_D$	$12.75 - 10.50 = 2.25 < 2.40 = R_2$	No significativa
$\mu_D - \mu_A$	$10.50 - 7.25 = 3.25^* > 2.60 = R_3$	Significativa
$\mu_D - \mu_B$	$10.50 - 8.50 = 2.0 < 2.40 = R_2$	No significativa
$\mu_B - \mu_A$	$8.50 - 7.25 = 1.25 < 2.40 = R_2$	No significativa

De esta tabla se concluye que $\mu_A = \mu_B$, $\mu_B = \mu_D$ y $\mu_C = \mu_D$, mientras que $\mu_A \neq \mu_C$, $\mu_B \neq \mu_C$ y $\mu_A \neq \mu_D$, que son las mismas conclusiones que se obtuvieron con el método LSD. En general, las pruebas de Duncan y LSD tienen un desempeño similar.

Tratamiento control
Se refiere a un tratamiento estándar de referencia o a la ausencia de tratamiento.

Comparación de tratamientos con un control (método de Dunnet)

Una vez que se rechaza H_0 con el ANOVA, en ocasiones uno de los k tratamientos a comparar es el llamado *tratamiento control* y el interés fundamental es comparar los

$k-1$ tratamientos restantes con dicho control. En muchos casos el tratamiento control se refiere a un tratamiento estándar de referencia o también a la ausencia de tratamiento (véase ejercicio 3.12). Por ejemplo, al comparar varios medicamentos para el resfriado es conveniente que uno de los tratamientos sea que los pacientes no utilicen ningún medicamento; esto sirve como referencia para decidir la posible utilidad de los medicamentos.

Por facilidad, denotemos como tratamiento control al k-ésimo tratamiento. Hacer comparaciones con respecto al control implica probar las $k-1$ hipótesis dadas por:

$$H_0 : \mu_i = \mu_k$$
$$H_A : \mu_i \neq \mu_k$$

con $i = 1, 2, \ldots, k-1$, donde k es el tratamiento control. La hipótesis nula se rechaza si,

$$\left|\bar{Y}_{i\bullet} - \bar{Y}_{k\bullet}\right| > D_\alpha(k-1, l)\sqrt{CM_E\left(\frac{1}{n_i}+\frac{1}{n_k}\right)}$$

donde $D_\alpha(k-1, l)$ se encuentra en las tablas del apéndice; l son los grados de libertad del cuadrado medio del error. Se recomienda que el tamaño de muestra del tratamiento control sea grande, a fin de estimar su media con mayor precisión.

Comparación por contrastes

No siempre interesa probar sólo las $k(k-1)/2$ hipótesis dos a dos dadas por $H_0 : \mu_i = \mu_j$ *vs.* $H_0 : \mu_i \neq \mu_j$ para $i \neq j$, y no siempre estas hipótesis dos a dos interesan todas por igual. En ocasiones, el objetivo del estudio lleva a contrastar hipótesis que involucran a más de dos medias. En esta sección se presentan este tipo de alternativas en la comparación de medias, pero antes se definen los conceptos de *contraste* y *contrastes ortogonales*.

Contraste

Una expresión de la forma $C = \Sigma_{i=1}^{k} c_i \mu_i$ es una combinación lineal de las medias poblacionales de interés, donde los coeficientes c_i son números reales. La combinación lineal C se llama *contraste* si cumple que la suma de los coeficientes es igual a cero ($\Sigma_{i=1}^{k} c_i = 0$). Muchas hipótesis estadísticas de interés son contrastes, como por ejemplo las hipótesis de comparación de medias. En efecto, ya hemos visto que la hipótesis nula $H_0 : \mu_i = \mu_j$ para $i \neq j$ se puede escribir de manera equivalente como $H_0 : \mu_i - \mu_j = 0$, donde se observa que el contraste correspondiente es la combinación lineal $c_i\mu_i + c_j\mu_j$ con $c_i = 1$ y $c_j = -1$, e interesa verificar si es estadísticamente igual a cero.

Contraste
Combinación lineal de medias poblacionales donde la suma de los coeficientes es igual a cero.

En general, supongamos que interesa probar si el contraste definido por $C = \Sigma_{i=1}^{k} c_i \mu_i$ es igual a cero. Si las poblaciones objeto de estudio son normales ($N(\mu_i, \sigma_i^2); i = 1, 2, \ldots, k$) el contraste C sigue una distribución normal con media $\mu_C = \Sigma_{i=1}^{k} c_i \mu_i$ y varianza $V_C = \Sigma_{i=1}^{k} \frac{c_i^2}{n_i}\sigma_i^2$. Cuando las varianzas de los tratamientos

son iguales y el diseño experimental es balanceado ($n_i = n$ para cada i), la varianza del contraste se reduce a $V_C = \frac{\sigma^2}{n}\Sigma_{i=1}^{k} c_i^2$. Al usar el CM_E para estimar a σ^2 y $\bar{Y}_{i\bullet}$ para estimar a la media μ_i, se puede ver que un intervalo al $100(1-\alpha)\%$ de confianza para el contraste C está dado por:

$$\sum_{i=1}^{k} c_i \bar{Y}_{i\bullet} \pm t_{\alpha/2,\, N-k} \sqrt{\frac{CM_E}{n} \sum_{i=1}^{k} c_i^2}$$

donde $t_{\alpha/2,\, N-k}$ es un punto porcentual de la distribución T de Student con $N-k$ grados de libertad. En caso de que el intervalo contenga al cero se concluye que el contraste C es estadísticamente igual a cero.

Contrastes ortogonales
Cuando la suma del producto de los coeficientes de dos contrastes es igual a cero.

Contrastes ortogonales

En el caso de un diseño balanceado, dos contrastes $C_1 = \Sigma_{i=1}^{k} c_{1i}\mu_i$ y $C_2 = \Sigma_{i=1}^{k} c_{2i}\mu_i$ son *ortogonales* si la suma del producto de los coeficientes es igual a cero, esto es, si $\Sigma_{i=1} c_{1i}c_{2i} = 0$; para el diseño desbalanceado son ortogonales si $\Sigma_{i=1} n_i c_{1i}c_{2i} = 0$. Dadas las k medias de interés correspondientes a k tratamientos objeto de estudio, se pueden construir una infinidad de conjuntos de $k-1$ contrastes ortogonales entre sí. En particular, con el uso de contrastes ortogonales es posible construir un grupo de hipótesis de interés independientes entre sí. Por ejemplo, en el problema de los $k = 4$ métodos de ensamble se pueden construir grupos de contrastes ortogonales de tamaño tres. Una posibilidad de elección se muestra en la siguiente tabla:

c_1	c_2	c_3	c_4	Contrastes ortogonales
2	–1	–1	0	$2\mu_A - \mu_B - \mu_C$
0	1	–1	0	$\mu_B - \mu_C$
1	1	1	–3	$\mu_A + \mu_B + \mu_C - 3\mu_D$

Es fácil ver que los tres contrastes definidos en esta tabla son ortogonales entre sí. Por ejemplo, el primero y el segundo son ortogonales porque $(2 \times 0) + (-1 \times 1) + (-1 \times -1) + (0 \times 0) = 0$, y lo mismo pasa con los otros dos posibles productos. Observe también que con cada contraste se puede definir una hipótesis estadística, como se hace en el siguiente método de Sheffé.

Método de Sheffé
Sirve para probar todos los contrastes de medias que pudieran ser de interés, en particular aquellos que involucran a más de dos medias.

Método de Sheffé

Este método está diseñado para probar todos los contrastes de medias que pudieran interesar al experimentador, sin el inconveniente de inflar por ello el error tipo I (detección de diferencias que no existen). Supongamos que interesa contrastar las hipótesis

$$H_0 : 2\mu_A = \mu_B + \mu_C$$
$$H_A : 2\mu_A \neq \mu_B + \mu_C \qquad \textbf{(3.14)}$$

donde la hipótesis nula se puede escribir alternativamente como $H_0 : 2\mu_A - \mu_B - \mu_C = 0$, lo cual implica que la hipótesis está definida por el contraste $C_0 = 2\mu_A - \mu_B - \mu_C$. De manera que el contraste estimado está dado por

$$\hat{C}_0 = 2\bar{Y}_A - \bar{Y}_B - \bar{Y}_C$$

y su varianza estimada es

$$V(\hat{C}_0) = CM_E \sum \frac{c_i^2}{n_i}$$

donde n_i es el número de mediciones en el tratamiento $i = A, B, C$. Intervalos simultáneos al $100(1-\alpha)\%$ de confianza para todos los contrastes tienen la forma

$$\hat{C} \pm \sqrt{(k-1)V(\hat{C})F_{\alpha,\, k-1,\, N-k}}$$

donde $\hat{C}$ representa la estimación de cualquier posible contraste y $F_{\alpha,\, k-1,\, N-k}$ es el cuantil $100(1-\alpha)$ de una distribución F con $k-1$ grados de libertad en el numerador, y $N-k$ grados de libertad en el denominador. Si el intervalo resultante para un contraste particular, digamos C_0, no contiene al cero, se concluye que el contraste es significativamente diferente de cero, lo cual lleva a rechazar H_0. De manera equivalente, el método de Sheffé rechaza la hipótesis nula si el contraste asociado es

$$\left|\hat{C}_0\right| > \sqrt{(k-1)V(\hat{C})F_{\alpha,\, k-1,\, N-k}}$$

Supongamos que en el ejemplo de los métodos de ensamble se quieren contrastar las hipótesis dadas en la ecuación (3.14). Con las medias muestrales (tabla 3.7) se calcula el estadístico $\hat{C}_0 = 2(7.25) - 8.50 - 12.75 = -6.75$. La varianza del contraste es $V(\hat{C}_0) = 2.46(6)/4 = 3.69$. Como $\sqrt{(k-1)V(\hat{C})F_{\alpha,\, k-1,\, N-k}} = \sqrt{3 \times 3.69 \times 3.49} = 6.21$ y $\left|\hat{C}_0\right| = 6.75$, se rechaza la hipótesis $H_0 : 2\mu_A = \mu_B + \mu_C$ y se acepta la $H_A : 2\mu_A \neq \mu_B + \mu_C$.

Verificación de los supuestos del modelo

La validez de los resultados obtenidos en cualquier análisis de varianza queda supeditado a que los supuestos del modelo se cumplan. Estos supuestos son: *normalidad*, *varianza constante* (*igual varianza de los tratamientos*) e *independencia*. Esto es, la respuesta (Y) se debe distribuir de manera normal, con la misma varianza en cada tratamiento y las mediciones deben ser independientes. Estos supuestos sobre Y se traducen en supuestos sobre el término error (ε) en el modelo [véase expresión (3.2)]. Es una práctica común utilizar la muestra de *residuos* para comprobar los supuestos del modelo, ya que si los supuestos se cumplen, los residuos o residuales se pueden ver como una muestra aleatoria de una distribución normal con media cero y varianza constante. Los residuos, e_{ij}, se definen como la diferencia entre la respuesta observada (Y_{ij}) y la respuesta predicha por el modelo ($\hat{Y}_{ij}$), lo cual permite hacer un diagnóstico más directo de la calidad del modelo, ya que su magnitud señala qué tan bien describe a los datos el modelo. Veamos.

Residuos
Son generados por la diferencia entre la respuesta observada y la respuesta predicha por el modelo en cada prueba experimental.

Recordemos de (3.2), que el modelo que se espera describa los datos en el DCA está dado por:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \tag{3.15}$$

donde Y_{ij} ($i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n$) es el j-ésimo dato en el tratamiento i; μ es la media global, τ_i es el efecto del tratamiento i y ε_{ij} representa al error asociado con la observación Y_{ij}. Cuando se realiza el ANOVA, y sólo cuando éste resulta significativo, entonces se procede a estimar el modelo ajustado o modelo de trabajo dado por:

$$\hat{Y}_{ij} = \hat{\mu} + \hat{\tau}_i \tag{3.16}$$

donde $\hat{Y}_{ij}$ es la respuesta predicha, $\hat{\mu}$ es la media global estimada y $\hat{\tau}_i$ es el efecto estimado del tratamiento i; los gorros indican que son *estimadores*, es decir, valores calculados a partir de los datos del experimento. El término del error desaparece del modelo estimado, por el hecho de que su valor esperado es igual a cero ($E(\varepsilon_{ij}) = 0$). Como la media global se estima con $\bar{Y}_{\bullet\bullet}$ y el efecto del tratamiento con $\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$, el modelo ajustado del DCA se puede escribir como:

$$\hat{Y}_{ij} = \bar{Y}_{\bullet\bullet} + (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}) = \bar{Y}_{i\bullet} \tag{3.17}$$

Esto es, la respuesta predicha para cada observación es la media muestral del tratamiento correspondiente. De esta manera, el residual o *residuo asociado a la observación* Y_{ij} está dado por

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i\bullet}$$

Los supuestos del modelo lineal (3.15), en términos de los residuos, son:

1. *Los* e_{ij} *siguen una distribución normal con media cero.*
2. *Los* e_{ij} *son independientes entre sí.*
3. *Los residuos de cada tratamiento tienen la misma varianza* σ^2.

Para comprobar cada supuesto existen pruebas analíticas y gráficas que veremos a continuación. Por sencillez, muchas veces se prefieren las pruebas gráficas. Éstas tienen el inconveniente de que no son "exactas", pero aun así, en la mayoría de las situaciones prácticas proporcionan la evidencia suficiente en contra o a favor de los supuestos. El uso de las pruebas gráficas requiere una fuerte evidencia visual para concluir que el supuesto en cuestión no se cumple, ya que se requiere que la evidencia en contra de un supuesto esté soportada por más de dos puntos. Cuando son uno o dos los puntos que se salen del comportamiento esperado de las gráficas se puede tratar de un problema de puntos aberrantes, no de violación del supuesto en cuestión. En ese caso debe investigarse la obtención de dichas mediciones atípicas, ya que ese tipo de puntos pueden afectar sensiblemente los resultados del análisis.

Se puede utilizar una prueba analítica para subsanar las ambigüedades que surjan en la interpretación visual (subjetiva) de las gráficas.

Es mejor prevenir en lo posible que los supuestos no se violen, para ello se aplican los tres principios básicos del diseño de experimentos: repetición, aleatorización y bloqueo. Es fácil encontrar situaciones en las que por no aplicar alguno de estos principios no se cumplen los supuestos del modelo. Por ejemplo, por no aleatorizar el orden en que se corren las pruebas pueden surgir problemas con el supuesto de independencia.

Normalidad

Gráfica de probabilidad
Sirve para verificar visualmente si los datos siguen una distribución de probabilidad específica.

Un procedimiento gráfico para verificar el cumplimiento del supuesto de normalidad de los residuos consiste en graficar los residuos en *papel* o en la *gráfica de probabilidad normal* que se incluye casi en todos los paquetes estadísticos. Esta gráfica del tipo *X-Y* tiene las escalas de tal manera que si los residuos siguen una distribución normal, al graficarlos tienden a quedar alineados en una línea recta; por lo tanto, si claramente no se alinean se concluye que el supuesto de normalidad no es correcto. Cabe enfatizar el hecho de que el ajuste de los puntos a una recta no tiene que ser perfecto, dado que el análisis de varianza resiste pequeñas y moderadas desviaciones al supuesto de normalidad. En las figuras 3.6*a* y 3.6*b* se representan, en la gráfica de probabilidad normal, dos aspectos de los residuos, en los cuales el supuesto de normalidad no se cumple.

Gráfica de probabilidad en papel normal

Consideremos los N residuos e_i que resultan del análisis de una varianza, o cualquier conjunto de N datos de los cuales se quiere verificar su procedencia de una distribución normal. Los pasos en la construcción de la gráfica de probabilidad normal para los residuos son los siguientes:

1. Ordenar los N valores del menor al mayor y asignarles los rangos de 1 a N. Sean r_i, $i = 1, 2,\ldots, N$, los datos en orden creciente.
2. Calcular una posición de graficación para cada dato en función de su rango y del total de observaciones como $(i - 0.5)/N$, $i = 1, 2,\ldots, N$.
3. El papel de probabilidad normal es un formato para realizar una gráfica del tipo *X-Y*, donde una de las escalas es lineal y la otra es logarítmica. Sobre el papel de probabilidad normal se dibujan las parejas $(r_i, (i - 0.5)/N)$.
4. Dibujar una línea recta sobre los puntos para tratar de dilucidar si se ajustan a ella o no. La interpretación de la gráfica es subjetiva, pero muchas veces es suficiente para llegar a una conclusión razonable sobre la distribución que siguen los datos.

Para ilustrar lo anterior, supongamos que los residuos son los siguientes 10 datos: 48.8, 51.5, 50.6, 46.5, 41.7, 39.9, 50.4, 43.9, 48.6, 48.6. Los cálculos necesarios para obtener las parejas a graficar se muestran en la tabla 3.9.

En el papel de probabilidad normal se grafican las parejas dadas por la primera y tercera columnas $(r_i, (i - 0.5)/N)$, y la gráfica resultante se muestra en la figura 3.5*a*. En ésta no hay evidencia suficiente en contra de la normalidad de los datos.

Tabla 3.9 Cálculos para realizar una gráfica de probabilidad normal.

Dato r_i	Rango i	$(i - 0.5)/N$	$Z_i = \Phi^{-1}((i - 0.5)/N)$
39.9	1	0.05	–1.64
41.7	2	0.15	–1.03
43.9	3	0.25	–0.67
46.5	4	0.35	–0.38
48.6	5	0.5	0.00
48.6	6	0.5	0.00
48.8	7	0.65	0.38
50.4	8	0.75	0.67
50.6	9	0.85	1.03
51.5	10	0.95	1.64

Gráfica de probabilidad normal en papel ordinario

A falta de papel de probabilidad normal, la gráfica de probabilidad también se puede hacer en papel ordinario con escalas equiespaciadas en ambos ejes. Para ello, primero se obtiene el valor normal estandarizado Z_i que cumple la relación:

$$\frac{(i-0.5)}{N} = P(Z < Z_i) = \Phi(Z_i) \qquad \textbf{(3.18)}$$

donde $\Phi(Z_i)$ es la función de distribución normal estándar acumulada evaluada en Z_i. Es decir, $Z_i = \Phi^{-1}(\frac{i-0.5}{N})$. Las parejas a dibujar en el papel ordinario son (r_i, Z_i) (ver tabla 3.9). En la figura 3.5*b* se muestra la gráfica de probabilidad en papel ordinario para los mismos datos graficados en papel normal. Observe que es básicamente la misma gráfica. Los cálculos necesarios para los Z_i se pueden hacer fácilmente en Excel con la función: DISTR.NORM.ESTAND.INV y en *Statgraphics* con la función INVNORMAL.

Además de la evaluación visual basada en la gráfica de probabilidad normal, existen varios métodos analíticos para contrastar la hipótesis H_0 : *Hay normalidad* contra H_A : *No hay normalidad*. Entre dichas pruebas se encuentran la ji-cuadrada para bondad de ajuste, la prueba de Shapiro-Wilks y la prueba de Anderson-Darling,

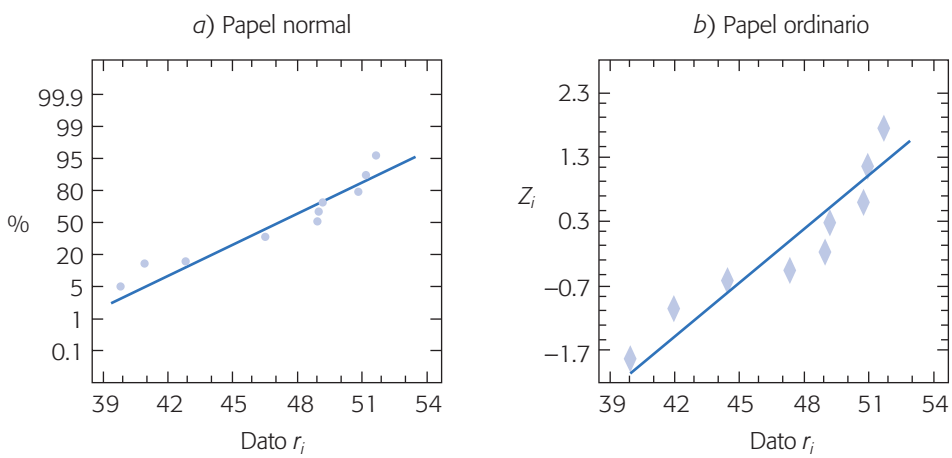


Figura 3.5 Gráfica de probabilidad en papel normal y en papel ordinario.

de las cuales, la de Shapiro-Wilks es una de las más recomendadas y que presentamos a continuación.

Prueba de Shapiro-Wilks para normalidad

Consideremos una muestra aleatoria de datos $x_1, x_2, \ldots, x_n$ que proceden de cierta distribución desconocida denotada por $F(x)$. Se quiere verificar si dichos datos fueron generados por un proceso normal, mediante las hipótesis estadísticas:

H_0 : *Los datos proceden de una distribución normal* ($F(x)$ *es normal*).
H_A : *Los datos no proceden de una distribución normal* ($F(x)$ *no es normal*).

Los pasos para la prueba de Shapiro-Wilks son: 1) Se ordenan los datos de menor a mayor. Denotemos los datos ordenados por $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. 2) De la tabla dada en el apéndice para este procedimiento se obtienen los coeficientes $a_1, a_2, \ldots, a_k$, donde k es aproximadamente $n/2$. 3) Se calcula el estadístico W definido como:

$$W = \frac{1}{(n-1)S^2}\left[\sum_{i=1}^{k} a_i (X_{(n-i+1)} - X_{(i)})\right]^2 \qquad \textbf{(3.19)}$$

donde S^2 es la varianza muestral. 4) Por último, si el valor del estadístico es mayor que su valor crítico al nivel α seleccionado en la tabla del apéndice, se rechaza la normalidad de los datos.

Para ilustrar la prueba de Shapiro-Wilks consideremos otra vez los mismos datos de las gráficas de probabilidad normal. De acuerdo con los datos ordenados, parte del procedimiento posterior al paso 2 para calcular el estadístico W se resume en la tabla que se presenta más adelante.

La varianza es $S^2 = 15.72$. Con la fórmula de la ecuación (3.19) se obtiene que

$$W = \frac{1}{(10-1)15.72}[11.26]^2 = 0.896$$

i	a_i	$(X_{(n-i+1)} - X_{(i)})$	$a_i(X_{(n-i+1)} - X_{(i)})$
1	0.5739	51.5 – 39.9 = 11.6	6.66
2	0.3291	50.6 – 41.7 = 8.9	2.93
3	0.2141	50.4 – 43.9 = 6.5	1.39
4	0.1224	48.8 – 46.5 = 2.3	0.28
5	0.0399	48.6 – 48.6 = 0	0.00

Con el tamaño de muestra $n = 10$, en la tabla de valores críticos dada en el apéndice se lee que el cuantil 95 es $W_{1-0.05} = 0.987$. Como W es menor que $W_{1-\alpha}$ se acepta que los datos proceden de una distribución normal, que concuerda con lo que se observó en las gráficas de probabilidad de la figura 3.5.

Varianza constante

Varianza constante
Supuesto del ANOVA que se cumple cuando los tratamientos tienen la misma varianza.

Una forma de verificar el supuesto de *varianza constante* (o que los tratamientos tienen la misma varianza) es graficando los predichos contra los residuos ($\hat{Y}_{ij}$ *vs.* e_i), por lo general $\hat{Y}_{ij}$ va en el eje horizontal y los residuos en el eje vertical. Si los puntos

en esta gráfica se distribuyen de manera aleatoria en una banda horizontal (sin ningún patrón claro y contundente), entonces es señal de que se cumple el supuesto de que los tratamientos tienen igual varianza. Por el contrario, si se distribuyen con algún patrón claro y contundente, como por ejemplo una forma de "corneta o embudo", entonces es señal de que no se está cumpliendo el supuesto de varianza constante (figura 3.6*c*). Un claro embudo en los residuales indicará que el error de pronóstico del modelo tiene una relación directa (positiva o negativa) con la magnitud del pronóstico (predicho).

Otra gráfica que ayuda a verificar el supuesto es la gráfica de niveles de factor contra residuos. En el eje *X* de esta gráfica se ponen los tratamientos o los niveles de un factor, y en el eje vertical se agregan los residuos correspondientes a cada tratamiento o nivel de factor. Si se cumple el supuesto de varianza constante, se espera que la amplitud de la dispersión de los puntos en cada nivel de factor tenderá a ser similar; y no se cumplirá el supuesto si hay diferencias fuertes en esta amplitud,

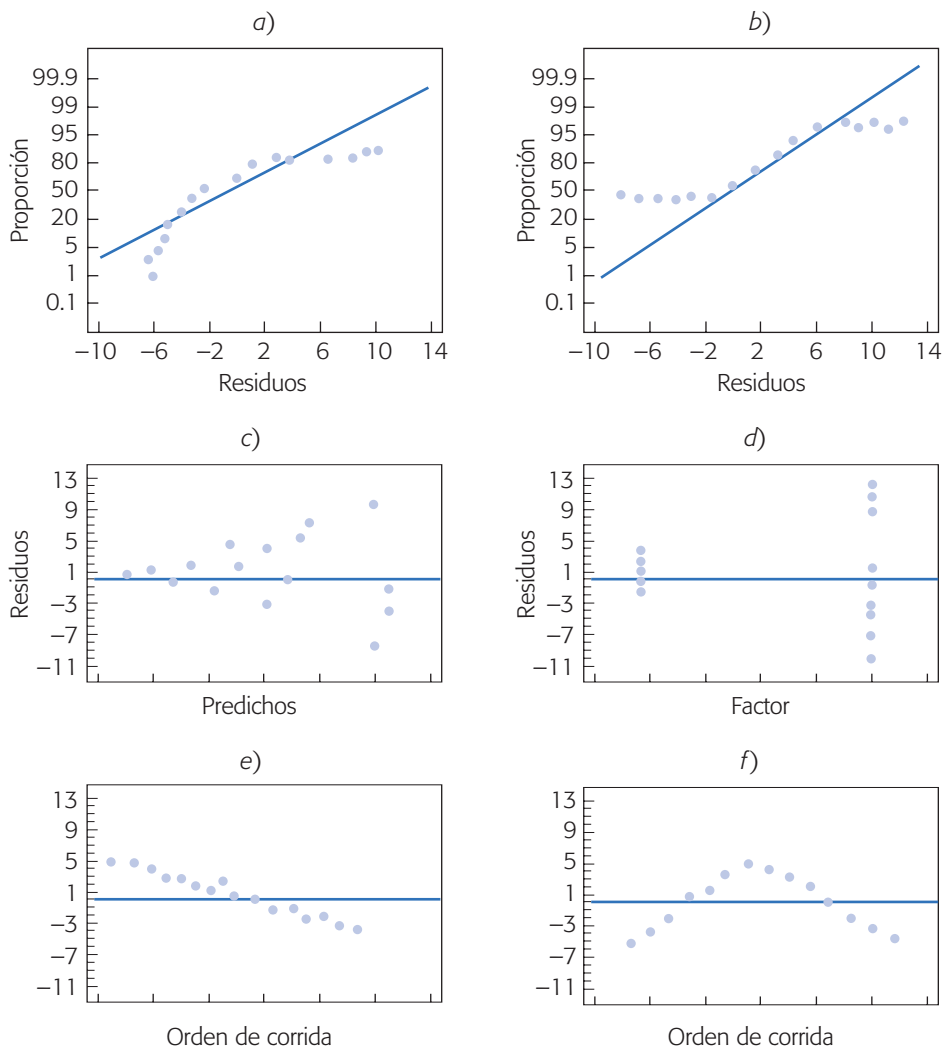


Figura 3.6 Ejemplos de gráficas de residuos donde no se cumplen los supuestos para el ANOVA.

como se muestra en la figura 3.6*d*. En la interpretación de esta gráfica debe considerarse que, en estadística, las pequeñas diferencias por lo general no son significativas, y también debe tomarse en cuenta la cantidad de observaciones hechas en cada nivel del factor, puesto que este hecho puede impactar la dispersión aparente en cada tratamiento.

Otra interpretación de la gráfica de factor contra residuos es que cuando los tratamientos o niveles muestran una dispersión diferente de sus residuales correspondientes (como en la figura 3.6*d*), es que el factor o los tratamientos tienen un efecto significativo sobre la variabilidad de la respuesta.

Con base en esta información se podría proponer un nivel de operación para dicho factor que minimice la dispersión y optimice la media.

Así, cuando hay una evidencia contundente en las gráficas anteriores, donde no se cumple el supuesto de varianza constante, entonces se debe ver en qué sentido resultan afectadas las conclusiones que se obtienen con el ANOVA y las pruebas de rangos múltiples. Por ejemplo, si se aprecia que el mejor tratamiento también es el que tiene menor dispersión, entonces se debe mantener tal tratamiento como la elección correcta, y ver si es de interés investigar por qué la diferencia en variabilidad con algunos de los otros tratamientos. Pero, si al que se le considera el mejor tratamiento es el que tiene la varianza más grande, entonces es difícil mantenerlo como la elección correcta. En este caso se debe replantear la decisión y el análisis. Una forma de volver a hacer el análisis y reconsiderar la situación es transformar los datos u observaciones Y_{ij}, de manera que se disminuyan las diferencias en dispersión y se pueda ver más claramente lo que ha pasado en el experimento. Existe una gran cantidad de transformaciones propuestas que logran lo anterior, entre las más frecuentes se encuentran la logarítmica y la raíz cuadrada. La transformación se hace de la siguiente manera: se saca logaritmo a los datos u observaciones por ejemplo, y con los datos transformados se vuelve a hacer el análisis completo. En la sección "Transformaciones para estabilizar varianzas" del capítulo 5 aborda el tema con detalle.

En general, siempre se debe investigar por qué no se ha cumplido el supuesto de varianza constante, ya que eso ayuda a entender mejor el proceso o sistema con el que se experimenta. Por ejemplo, una razón frecuente que hace que tal supuesto no se cumpla es que algunas variables tienen una dispersión directamente proporcional a su magnitud, de tal forma que si sus valores son pequeños, éstos tienden a ser más homogéneos en comparación con la variabilidad que entre sí tienen los valores grandes. Ahora veamos una prueba analítica para la igualdad de varianzas.

Prueba de Bartlett para homogeneidad de varianzas

Supongamos que se tienen k poblaciones o tratamientos independientes, cada uno con distribución normal ($N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, k$), donde las varianzas son desconocidas. Se quiere probar la hipótesis de igualdad de varianzas dada por:

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2 \\ H_A &: \sigma_i^2 \neq \sigma_j^2 \text{ para algún } i \neq j \end{aligned} \qquad \textbf{(3.20)}$$

Mediante un diseño completamente al azar se obtienen k muestras aleatorias de tamaños n_i ($i = 1, 2, \ldots, k$) de dichas poblaciones, de modo que el total de mediciones

es $N = n_1 + n_2 + \ldots + n_k$. El estadístico de prueba para la hipótesis (3.20) está dado por

$$\chi_0^2 = 2.3026\frac{q}{c}$$

donde

$$q = (N-k)\log_{10} S_p^2 - \sum_{i=1}^{k}(n_i - 1)\log_{10} S_i^2$$

y

$$c = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}(n_i - 1)^{-1} - (N-k)^{-1}\right)$$

con

$$S_p^2 = \frac{\sum_{i=1}^{k}(n_i - 1)S_i^2}{N-k}$$

donde S_i^2 es la varianza muestral del tratamiento i. Bajo la hipótesis nula de igualdad de varianza, el estadístico χ_0^2 sigue una distribución ji-cuadrada con $k - 1$ grados de libertad, por lo que se rechaza H_0 cuando χ_0^2 es más grande que $\chi^2_{(\alpha,\, k-1)}$. Observe que el estadístico q, en el numerador del estadístico χ_0^2, es grande en la medida de que las varianzas muestrales S_i^2 son diferentes y es igual a cero cuando éstas son iguales.

La prueba de Bartlett que acabamos de describir es sensible a la falta de normalidad de las poblaciones de interés, por lo que debe comprobarse el cumplimiento de este supuesto.

Independencia

La suposición de independencia en los residuos puede verificarse si se grafica el orden en que se colectó un dato contra el residuo correspondiente. De esta manera, si al graficar en el eje horizontal el tiempo (orden de corrida) y en el eje vertical los residuos, se detecta una tendencia o patrón no aleatorio claramente definido, esto es evidencia de que existe una correlación entre los errores y, por lo tanto, el supuesto de independencia no se cumple (véanse figuras 3.6*e* y 3.6*f*). Si el comportamiento de los puntos es aleatorio dentro de una banda horizontal, el supuesto se está cumpliendo. La violación de este supuesto generalmente indica deficiencias en la planeación y ejecución del experimento; asimismo, puede ser un indicador de que no se aplicó en forma correcta el principio de aleatorización, o de que conforme se fueron realizando las pruebas experimentales aparecieron factores que afectaron la respuesta observada. Por ello, en caso de tener problemas con este supuesto, las conclusiones que se obtienen del análisis son endebles y por ello es mejor revisar lo hecho y tratar de investigar por qué no se cumplió con ese supuesto de independencia, a fin de reconsiderar la situación.

Una prueba analítica para verificar la independencia entre residuos consecutivos es la prueba de Durbin-Watson, que se presenta en el capítulo 11. El problema con dicha prueba es que no es capaz de detectar otros patrones de correlación entre residuos (no consecutivos) que también son violatorios del supuesto de independencia.

Tabla 3.10 Residuos para ejemplo 3.2.

Cuero	Observado Y_{ij}	Predicho $\bar{Y}_{i\cdot}$	Residuo $e_{ij} = Y_{ij} - \bar{Y}_{i\cdot}$	Cuero	Observado Y_{ij}	Predicho $\bar{Y}_{i\cdot}$	Residuo $e_{ij} = Y_{ij} - \bar{Y}_{i\cdot}$
A	264	256.7	7.33	*A*	262	256.7	5.33
C	220	230.8	–10.83	*D*	220	220.7	–0.67
B	208	209.8	–2.5	*A*	255	256.7	–1.67
B	220	209.8	9.5	*B*	200	209.8	–10.5
A	260	256.7	3.33	*D*	222	220.7	1.33
A	258	256.7	1.33	*B*	213	209.8	2.5
D	217	220.7	–3.67	*A*	241	256.7	–15.67
C	263	230.8	32.17	*C*	228	230.8	–2.83
D	229	220.7	5.83	*B*	206	209.8	–4.5
C	219	230.8	–11.83	*C*	230	230.8	–0.83
B	216	209.8	5.5	*D*	215	220.7	–5.67
C	225	230.8	–5.83	*D*	224	220.7	3.33

Ejemplo 3.6

(Continuación del análisis para comparar cuatro tipos de cuero). En el ejemplo 3.2 se compararon cuatro tipos de cuero en cuanto a su desgaste, y mediante el ANOVA se concluyó que los cueros tienen un desgaste promedio diferente (ver tabla 3.5). Falta ver que se cumplan los supuestos del ANOVA. Para ello, primero se calculan los residuos de las 24 mediciones, restando a cada valor observado su correspondiente predicho, que en este caso como $\hat{Y}_{ij} = \bar{Y}_{i\cdot}$ se debe restar la media del tratamiento correspondiente. Los 24 residuos se listan en la tabla 3.10.

Con la muestra de 24 residuos se procede a dibujar las gráficas de residuos en papel de probabilidad normal, residuos contra predichos y residuos contra orden de corrida. Las gráficas resultantes se muestran en las figuras 3.7*a*, *b* y *c*. Se observa el cumplimiento de los supuestos de normalidad, varianza constante e independencia, respectivamente. Sin embargo, en las tres gráficas es notorio un punto que se aleja bastante del resto, el cual es un punto aberrante cuyo origen debe investigarse. En la tabla 3.10 se encuentra que este residuo grande de valor 32.17 y que corresponde a la prueba 8 con una medición de 263 en el tipo de cuero *C*. Debe verificarse que no haya ningún error con este dato. Cuando un punto aberrante no se percibe, puede afectar sensiblemente las conclusiones del análisis del experimento.

Elección del tamaño de la muestra

Una decisión importante en cualquier diseño de experimentos es decidir el número de réplicas que se hará por cada tratamiento (tamaño de muestra). Por lo general, si se esperan diferencias pequeñas entre tratamientos será necesario un mayor tamaño de muestra. Aunque existen varios métodos para estimar el tamaño muestral, muchas veces tienen poca aplicabilidad porque requieren cierto conocimiento previo sobre la varianza del error experimental.

Si recurrimos a la experiencia vemos que el número de réplicas en la mayoría de las situaciones experimentales en las que se involucra un factor varía entre cinco

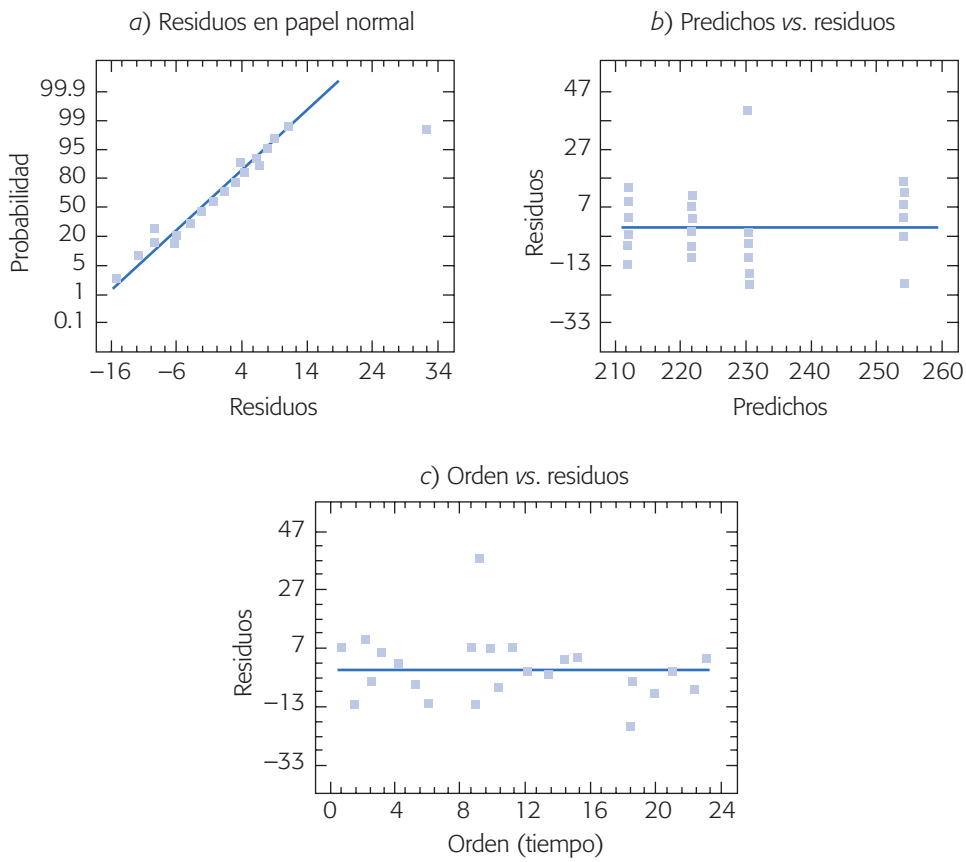


Figura 3.7 Gráficas de residuos para los tipos de cuero.

y diez; incluso, en algunos casos puede llegar hasta 30. La tendencia podría inclinarse por un extremo de este rango e incluso salirse de éste, de acuerdo con las siguientes consideraciones:

- A menor diferencia que se espera en los tratamientos, mayor será la cantidad de réplicas si se quieren detectar diferencias significativas, y viceversa, es decir, si se esperan grandes diferencias quizá con pocas réplicas sea suficiente.
- Si se espera mucha variación dentro de cada tratamiento, debido a la variación de fuentes no controladas como métodos de medición, medio ambiente, materia prima, etc., entonces se necesitarán más réplicas.
- Si son varios tratamientos (cuatro o más), entonces éste es un punto favorable para reducir el número de réplicas.

Además de lo anterior, es preciso considerar los costos y el tiempo global del experimento. De aquí que si se toman en cuenta las consideraciones antes expuestas se podrá establecer el tamaño de muestra que permita responder en una primera fase las preguntas más importantes que se plantearon con el experimento.

Elección del tamaño de muestra por intervalo de confianza

Supongamos que el experimentador ya tiene el número de tratamientos que desea probar, k, y que tomando en cuenta las consideraciones antes citadas tiene una propuesta inicial del número de réplicas por tratamiento que va a utilizar, n_0. También tiene una idea aproximada del valor de σ (la desviación estándar del error aleatorio), así como una idea de la magnitud de las diferencias, d_T, entre tratamientos que le interesa detectar. Por ejemplo, supongamos que en el caso de los tiempos promedio de los $k = 4$ métodos de ensamble (ejemplo 3.1), tiene idea de realizar $n_0 = 5$ pruebas; en cuanto a las diferencias, le interesa detectar 2 minutos, $d_T = 2$ entre un método y otro, y espera que cada método tenga una variabilidad intrínseca de $\sigma = 1.5$; esto debido a factores no controlados (habilidad del operador, cansancio, variabilidad de las partes a ensamblar, error de medición del tiempo de ensamble, etcétera).

Ahora recordemos que en las comparaciones o pruebas de rangos múltiples, la diferencia mínima significativa entre tratamientos está dada por la expresión (3.12):

$$LSD = t_{(\alpha/2,\, N-k)}\sqrt{2CM_E/n}$$

despejando n de aquí, obtenemos:

$$n = \frac{2\left(t_{(\alpha/2,\, N-k)}\right)^2 CM_E}{(LSD)^2}$$

Si la significancia es $\alpha = 0.05$, entonces en esta fórmula se hacen las siguientes sustituciones: $N = k \times n_0$, $CM_E = \sigma^2$, $LSD = d_T$; de esta forma, el tamaño de muestra que tentativamente se debe usar está dado por,

$$n = \frac{2\left(t_{(0.025,\, k\times n_0 - k)}\right)^2 \sigma^2}{(d_T)^2}$$

El valor de n arrojado por esta fórmula dará una idea del número de réplicas por tratamiento, de acuerdo con las consideraciones iniciales que se reflejan a través de (k, n_0, σ, d_T), y sobre todo por el número total de corridas experimentales, $N = k \times n$, que es lo que muchas veces interesa más al experimentador debido a los costos y tiempos. Si N está fuera del presupuesto se podrán revisar algunas consideraciones y quizá pensar en un número menor de tratamientos.

Al aplicar esta expresión al caso de los cuatro métodos de ensamble obtenemos:

$$n = \frac{2(t_{(0.025,\,15)})^2\,(1.5)^2}{(2)^2} = \frac{2(2.131)^2\,(1.5)^2}{(2)^2} = 5.1$$

Por lo tanto, $n = 5$ se debería utilizar como tamaño de muestra (número de pruebas por tratamiento).

Uso de software computacional

Casi cualquier *software estadístico* incluye procedimientos para realizar análisis de varianza, comparar tratamientos y hacer análisis relacionados. En términos generales,

en una columna se registra el código para cada tratamiento corrido (se ponen tantos renglones como pruebas hechas), y en otra columna se registran los valores correspondientes obtenidos para *Y*. Con esto, en *Statgraphics* el análisis de los diseños comparativos se realiza básicamente en la opción *Compare* del menú principal.

La secuencia para un diseño completamente al azar es: *Compare* → *Analysis of variance* → *One-way anova*. En las opciones del procedimiento aparecen todas las pruebas y análisis que se han descrito en este capítulo.

Otra posibilidad en *Statgraphics* es accesar con la siguiente secuencia de opciones: *Special* → *Experimental Design* → *Create Design*, después de esto se debe elegir el tipo de diseño, que en este caso es *Single Factor Categorical*. Enseguida se define el número de niveles (tratamientos) y el nombre de los mismos. También se debe definir el nombre de la(s) variable(s) de respuesta(s). En la siguiente pantalla se pedirá el número de réplicas adicionales a la básica (si se pide una, en total se tendrán dos al considerar la réplica básica) y también aparece la opción de aleatorizar el orden para correr las pruebas, que siempre debe utilizarse en un diseño completamente aleatorizado. Todo esto permitirá generar una columna en la que se incluyen todas las pruebas a ser corridas, y una columna en blanco para cada variable de respuesta, la cual debe ser llenada en la medida que se vayan obteniendo los resultados del experimento. De la versión 15 de *Statgraphics* en adelante, la secuencia para crear diseños es *DOE* → *Design Creation*.

Para hacer el análisis, una vez generado el archivo de datos con los tratamientos y las respuestas, se siguen las opciones: *Special* → *Experimental Design* → *Analyze Design*, después se da el nombre de la variable de respuesta a analizar, y entonces se tendrá acceso a un conjunto de opciones de análisis tanto gráficas como analíticas, entre ellas las que hemos comentado en este capítulo.

En Minitab se registran los datos en dos columnas, como ya se dijo, y al ANOVA se accesa con la secuencia *Stat* → *Anova* → *One way*, y se da el nombre de las columnas que contienen los datos. También se eligen las comparaciones de medias deseadas y las gráficas.

Uso de Excel

El ANOVA de un diseño con un criterio de clasificación se realiza con la secuencia: *Herramientas* → *Análisis de datos* → *Análisis de varianza con un factor*. Si no estuviera activada la opción de *Análisis de datos*, se utiliza la opción de *Complementos* dentro del mismo menú de *Herramientas*. Entonces, se declara el rango de los datos, que pueden estar acomodados por columnas o por renglones. La salida contiene las estadísticas básicas de cada una de las muestras y el ANOVA correspondiente.

Preguntas y ejercicios

1. Explique en qué consiste y cuándo se debe aplicar el diseño completamente al azar con un solo criterio de clasificación.
2. Supongamos que se desea probar la igualdad entre sí de cinco medias. Una alternativa para hacer esto sería comparar de dos en dos las medias, utilizando la prueba *T* de Student y al final tomar una decisión. Explique por qué esto aumenta el error tipo I.

3. ¿Qué mide el cuadrado medio del error en el ANOVA de un experimento?
4. ¿Qué son los grados de libertad para una suma de cuadrados en un análisis de varianza?
5. A continuación se muestra parte del ANOVA para comparar cinco tratamientos con cuatro réplicas cada uno.

Fuente de variación	Suma de cuadrados	G. de libertad	C. medio	Razón *F*	Valor-*p*
Tratamiento	800				
Error	400				
Total					

   a) Agregar en esta tabla los grados de libertad, el cuadrado medio y la razón *F* para cada una de las fuentes de variación.
   b) Explique de manera esquemática cómo calcularía el valor-*p* o la significancia observada, para ver si hay diferencia entre tratamientos.
   c) ¿Con la información disponible se pueden hacer conjeturas sobre si hay diferencias significativas entre tratamientos? Argumente su respuesta.
   d) Anote el modelo estadístico y formule la hipótesis pertinente.
6. Se desea investigar el efecto del pH en el crecimiento de cierto microorganismo en un medio específico. Para ello se realiza un experimento, teniendo como punto de partida la misma cantidad de microorganismos. Se hacen cuatro repeticiones y se obtienen los siguientes resultados. ¿Estos datos son evidencia suficiente para afirmar que los niveles de pH donde se logra menor y mayor crecimiento son e1 3 y el 2, respectivamente? Explique su respuesta.

Nivel de pH	Crecimiento promedio (en %)
1	80
2	105
3	75

7. Se desea investigar la influencia de la temperatura en el rendimiento de un proceso químico, en particular interesa investigar un rango de temperatura entre 60 y 120°C. Se tienen recursos para realizar 20 corridas experimentales.
   a) Los niveles de temperatura con los que se experimenta son: 60, 65, 70 y 120; se hacen cinco repeticiones con cada nivel. ¿Considera que es adecuado el diseño experimental usado? Argumente su respuesta, y de ser necesario proponga alternativas.
   b) El orden en que decidieron hacer las corridas experimentales para facilitar el trabajo experimental fue: primero las cinco del nivel bajo de temperatura, luego las cinco del siguiente y así hasta finalizar. ¿Es correcto lo que hicieron? Argumente su respuesta.
   c) Para hacer el análisis estadístico se comparan, mediante una prueba *T* de Student, de dos en dos niveles de temperatura, y con base en esto obtuvieron conclusiones. ¿Es adecuado tal análisis?, argumente, y en su caso proponga alternativas.
8. Describa en qué consiste cada uno de los supuestos del modelo en un análisis de varianza, y explique la forma típica en que estos supuestos se verifican.
9. ¿Qué son y cuándo se aplican las pruebas para comparar medias?

10. En una industria química se prueban diferentes mezclas para ver si difieren en cuanto al peso molecular final. Se prueban cuatro diferentes mezclas, con cinco repeticiones cada una. A continuación se muestra una parte de la tabla del análisis de varianza y los promedios obtenidos para cada mezcla.

Fuente de variación	Valor *p*
Mezcla	0.01
Error	

Mezcla	Peso medio
A	10 000
B	7 000
C	8 000
D	7 500

a) ¿Las mezclas difieren de manera significativa en cuanto a su peso molecular?
b) Con el análisis de varianza y de acuerdo al promedio, ¿se puede asegurar que con la mezcla *B* se logra un menor peso molecular? Argumente su respuesta.
c) Si al verificar los supuestos de varianza constante (igual varianza entre las mezclas), éstos no se cumplen, ¿qué significa eso? ¿Se puede seguir apoyando la conclusión del inciso *a*)?

11. Se hace un estudio sobre la efectividad de tres marcas de spray para matar moscas. Para ello, cada producto se aplica a un grupo de 100 moscas, y se cuenta el número de moscas muertas expresado en porcentajes. Se hacen seis réplicas y los resultados obtenidos se muestran a continuación.

	Número de réplica					
Marca de spray	1	2	3	4	5	6
1	72	65	67	75	62	73
2	55	59	68	70	53	50
3	64	74	61	58	51	69

a) Formule la hipótesis adecuada y el modelo estadístico.
b) ¿Existe diferencia entre la efectividad promedio de los productos en spray?
c) ¿Hay algún spray mejor? Argumente su respuesta.
d) Dé un intervalo al 95% de confianza para la efectividad promedio (porcentaje) de cada una de las marcas.
e) Dibuje las gráficas de medias y los diagramas de caja simultáneos, después interprételos.
f) Verifique los supuestos de normalidad y de igual varianza entre las marcas.

12. En un centro de investigación se realiza un estudio para comparar varios tratamientos que, al aplicarse previamente a los frijoles crudos, reducen su tiempo de cocción. Estos tratamientos son a base de bicarbonato de sodio ($NaHCO_3$) y cloruro de sodio o sal común (NaCl). El primer tratamiento es el de control, que consiste en no aplicar ningún tratamiento. El tratamiento T_2 es el remojo en agua con bicarbonato de sodio, el T_3 es remojar en agua con sal común y el T_4 es remojar en agua con una combinación de ambos ingredientes en proporciones iguales. La variable de respuesta es el tiempo de cocción en minutos. Los datos se muestran en la siguiente tabla:

Control	T_2	T_3	T_4
213	76	57	84
214	85	67	82
204	74	55	85
208	78	64	92
212	82	61	87
200	75	63	79
207	82	63	90

a) ¿De qué manera el experimentador debe aleatorizar los experimentos y el material experimental?
b) Dé ejemplos de factores que deben estar fijos durante las pruebas experimentales, para que no afecten los resultados y las conclusiones.
c) Formule y pruebe la hipótesis de que las medias de los tratamientos son iguales.
d) Obtenga el diagrama de caja y el gráfico de medias, después interprételos.
e) ¿Hay algún tratamiento mejor? ¿Cuál es el tiempo de cocción esperado para el mejor tratamiento?
f) Algo importante a cuidar en un experimento es que no haya efectos colaterales no deseados, causados por el tratamiento ganador; en este caso, piense en los posibles efectos colaterales que podría causar el mejor tratamiento.
g) ¿Se cumplen los supuestos del modelo? Verifique gráficamente.
h) Pruebe la hipótesis de igualdad de varianzas entre tratamientos (que corresponde a un supuesto).

13. Para estudiar la confiabilidad de ciertos tableros electrónicos para carros, se someten a un envejecimiento acelerado durante 100 horas a determinada temperatura, y como variable de interés se mide la intensidad de corriente que circula entre dos puntos, cuyos valores aumentan con el deterioro. Se probaron 20 módulos repartidos de manera equitativamente en cinco temperaturas y los resultados obtenidos fueron los siguientes:

20°C	40°C	60°C	80°C	100°C
15	17	23	28	45
18	21	19	32	51
13	11	25	34	57
12	16	22	31	48

a) Formule la hipótesis y el modelo estadístico para el problema.
b) Realice el análisis de varianza para estos datos, a fin de estudiar si la temperatura afecta la intensidad de corriente promedio.
c) ¿La temperatura afecta la variabilidad de las intensidades? Es decir, verifique si hay igual varianza entre los diferentes tratamientos.

14. En una empresa de manufactura se propone un tratamiento para reducir el porcentaje de productos defectuosos. Para validar esta propuesta se diseñó un experimento en el que se producía con o sin la propuesta de mejora. Cada corrida experimental consistió en producir un lote y la variable de respuesta es el porcentaje de producto defectuoso. Se hicieron 25 réplicas para cada tratamiento. Los datos obtenidos se muestran a continuación:

	Porcentaje de producto defectuoso												
Con tratam.	5.3	4.0	4.0	4.0	2.6	2.1	5.1	4.1	4.1	3.2	5.1	2.2	4.1
	2.2	1.1	2.0	3.0	3.1	2.1	1.2	3.3	2.1	4.0	2.0	3.0	
Sin tratam.	8.0	13.2	7.2	8.2	9.1	6.7	12.2	16.3	9.2	6.4	7.2	17.2	12.3
	8.7	11.3	4.5	6.6	9.2	10.2	10.6	13.3	5.2	6.2	8.0	4.8	

a) ¿Las diferencias son significativas estadísticamente?
b) ¿Cuál es el porcentaje de defectos que se espera con el nuevo tratamiento?
c) Cuantifique el nivel de reducción que se logró con el tratamiento propuesto.

15. Una compañía farmacéutica desea evaluar el efecto que tiene la cantidad de almidón en la dureza de las tabletas. Se decidió producir lotes con una cantidad determinada de almidón, y que las cantidades de almidón a aprobar fueran 2%, 5% y 10%. La variable de respuesta sería el promedio de la dureza de 20 tabletas de cada lote. Se hicieron 4 réplicas por tratamiento y se obtuvieron los siguientes resultados:

% de almidón	Dureza			
2	4.3	5.2	4.8	4.5
5	6.5	7.3	6.9	6.1
10	9.0	7.8	8.5	8.1

a) ¿Hay evidencia suficiente de que el almidón influye en la dureza en las tabletas? Halle el ANOVA.
b) Realice los análisis complementarios necesarios.
c) Si se desea maximizar la dureza de las tabletas, ¿qué recomendaría al fabricante?
d) Verifique los supuestos.

16. Los datos que se presentan enseguida son rendimientos en toneladas por hectárea de un pasto con tres niveles de fertilización nitrogenada. El diseño fue completamente aleatorizado, con cinco repeticiones por tratamiento.

Niveles de nitrógeno		
1	2	3
14.823	25.151	32.605
14.676	25.401	32.460
14.720	25.131	32.256
14.5141	25.031	32.669
15.065	25.267	32.111

a) ¿Las diferencias muestrales hacen obvia la presencia de diferencias poblacionales?
b) Obtenga el análisis de varianza e interprételo.
c) Analice los residuos, ¿hay algún problema?

17. Un químico del departamento de desarrollo de un laboratorio farmacéutico desea conocer cómo influye el tipo de aglutinante utilizado en tabletas de ampicilina de 500 mg en el porcentaje de friabilidad; para ello, se eligen los siguientes aglutinantes: polivinilpirrolidona (PVP), carboximetilcelulosa sódica (CMC) y grenetina (Gre). Los resultados del diseño experimental son los siguientes.

Aglutinante	% de friabilidad				
PVP	0.485	0.250	0.073	0.205	0.161
CMC	9.64	9.37	9.53	9.86	9.79
Gre	0.289	0.275	0.612	0.152	0.137

a) Especifique el nombre del diseño experimental.
b) ¿Sospecha que hay algún efecto significativo del tipo de aglutinante sobre la variable de respuesta?
c) Escriba las hipótesis para probar la igualdad de medias y el modelo estadístico.
d) Realice el análisis adecuado para probar las hipótesis e interprete los resultados.
e) Revise los supuestos, ¿hay algún problema?

18. Se cultivaron cuatro diferentes clonas de *agave tequilana* bajo un mismo esquema de manejo. Se quiere saber qué clona es la que responde mejor a dicho manejo, evaluando el nivel de respuesta con el porcentaje de azúcares reductores totales en base húmeda. Los datos se muestran a continuación:

Clona			
1	2	3	4
8.69	8.00	17.39	10.37
6.68	16.41	13.73	9.16
6.83	12.43	15.62	8.13
6.43	10.99	17.05	4.40
10.30	15.53	15.42	10.38

a) Mediante ANOVA, compare las medias de las clonas y verifique residuales.
b) ¿Hay una clona que haya respondido mejor al esquema de manejo? Argumente su respuesta.
c) En caso de que exista un empate estadístico entre dos o más clonas, ¿qué propondría para desempatar?

19. Uno de los defectos que causan mayor desperdicio en la manufactura de discos ópticos compactos son los llamados "cometas". Típicamente, se trata de una partícula que opone resistencia al fluido en la etapa de entintado. Se quiere comprobar de manera experimental la efectividad de un tratamiento de limpieza de partículas que está basado en fuerza centrípeta y aire ionizado. A 12 lotes de 50 CD se les aplica el tratamiento y a otros 12 lotes no se les aplica; en cada caso se mide el porcentaje de discos que presentan cometas, los resultados son los siguientes:

Con tratamiento	Sin tratamiento
5.30	8.02
4.03	13.18
4.03	7.15
4.00	8.23
2.56	9.11
2.05	6.66
5.06	12.15
4.06	16.3
2.08	9.20
4.03	6.35
2.04	7.15
1.18	8.66

a) Con el ANOVA vea si es efectivo el tratamiento de limpieza. ¿Debería implementarse?
b) ¿Es razonable suponer en el inciso *a*) que las varianzas son iguales?
c) ¿En qué porcentaje se reducen los discos con cometas?

Capítulo 4
Diseños de bloques

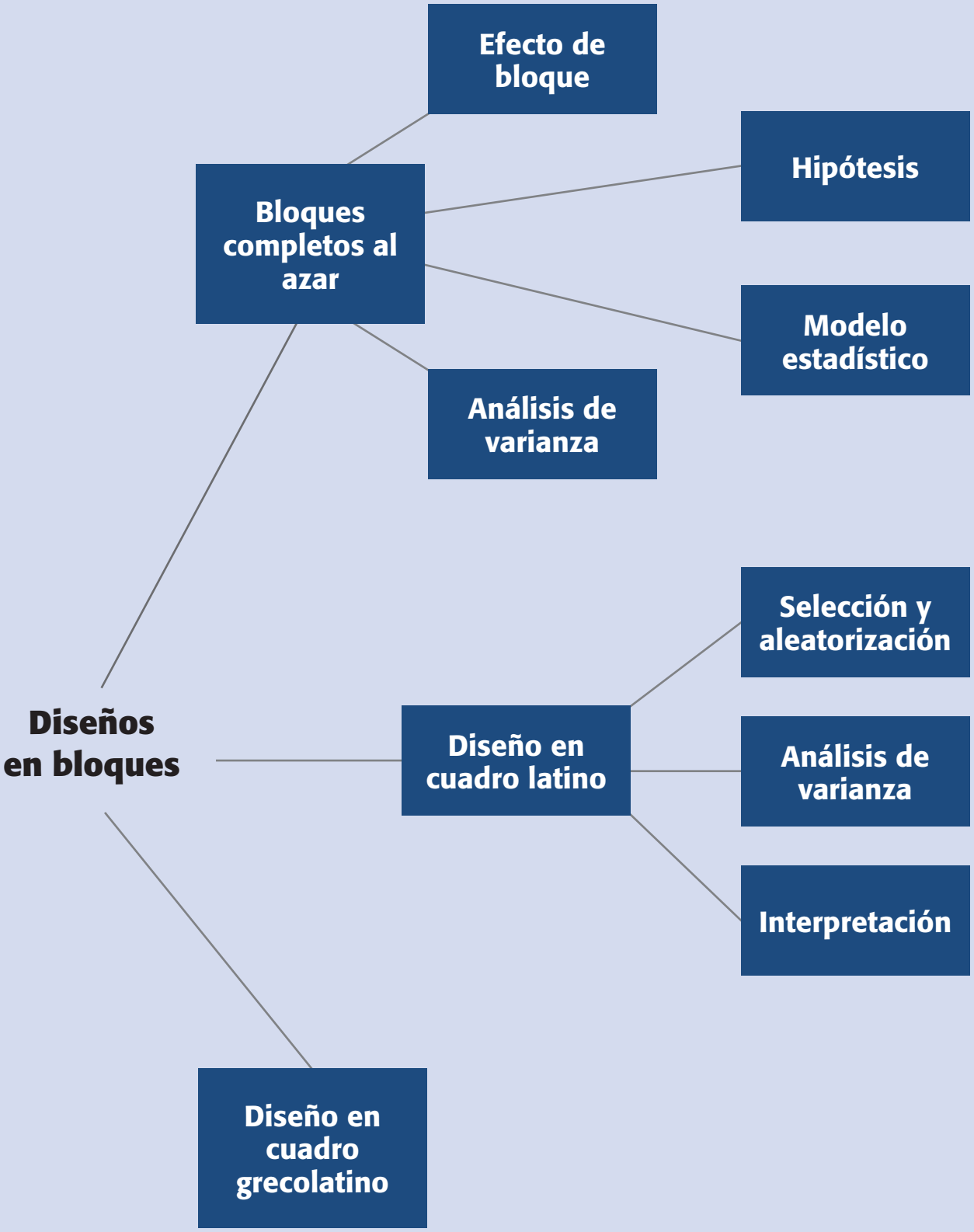
Sumario

- Diseños en bloques completos al azar
- Diseño en cuadro latino
- Diseño en cuadro grecolatino
- Uso de software

Objetivos de aprendizaje

- Identificar las características generales y los usos que se le dan a los diseños en bloques.
- Explicar la definición del diseño en bloques completo o al azar, así como su hipótesis, modelo estadístico y análisis de varianza.
- Describir la selección y la aleatorización del diseño en cuadro latino y su diferencia con el diseño en cuadro grecolatino.

Mapa conceptual



Conceptos clave
- Bloque completo
- Cuadro grecolatino
- Cuadro latino
- Cuadro latino estándar
- Efecto de interacción
- Fuentes de variabilidad

Diseño de bloques completos al azar

Cuando se quieren comparar ciertos tratamientos o estudiar el efecto de un factor, es deseable que las posibles diferencias se deban principalmente al factor de interés y no a otros factores que no se consideran en el estudio. Cuando esto no ocurre y existen otros factores que no se controlan o nulifican para hacer la comparación, las conclusiones podrían ser afectadas sensiblemente. Por ejemplo, supongamos que se quieren comprar varias máquinas, si cada máquina es manejada por un operador diferente y se sabe que éste tiene una influencia en el resultado, entonces es claro que el factor operador debe tomarse en cuenta si se quiere comparar a las máquinas de manera justa. Un operador más hábil puede hacer ver a su máquina (aunque ésta sea la peor) como la que tiene el mejor desempeño, lo cual impide hacer una comparación adecuada de los equipos. Para evitar este sesgo hay dos maneras de anular el posible efecto del factor operador: la manera lógica es utilizar el mismo operador en las cuatro máquinas; sin embargo, tal estrategia no siempre es aconsejable, ya que utilizar al mismo sujeto elimina el efecto del factor operador pero restringe la validez de la comparación con dicho operador, y es posible que el resultado no se mantenga al utilizar a otros operadores. La otra forma de anular el efecto operador en la comparación consiste en que cada operador trabaje durante el experimento con cada una de las máquinas. Esta estrategia es la más recomendable, ya que utilizar a todos los operadores con todas las máquinas permite tener resultados de la comparación que son válidos para todos los operadores. Esta última forma de nulificar el efecto de operadores, recibe el nombre de *bloqueo.*

Factores de bloque

Factores de bloque
Son las variables adicionales al factor de interés que se incorporan de manera explícita en un experimento comparativo para no sesgar la comparación.

A los factores adicionales al factor de interés que se incorporan de manera explícita en un experimento comparativo se les llama *factores de bloque.* Éstos tienen la particularidad de que no se incluyen en el experimento porque interese analizar su efecto, sino como un medio para estudiar de manera adecuada y eficaz al factor de interés. Los factores de bloque entran al estudio en un nivel de importancia secundaria con respecto al factor de interés y, en este sentido, se puede afirmar que se estudia un solo factor, porque es uno el factor de interés. Por ejemplo, en el caso de comparar cuatro máquinas que son manejadas por cuatro operadores, es pertinente incluir explícitamente al factor operadores (bloques) para lograr el propósito del estudio, pero esta inclusión no es con el fin de estudiar el efecto del factor operador (o comparar a los operadores). Más bien, la inclusión de los operadores es un medio y no un fin para lograr una comparación adecuada y eficaz de las máquinas.

Puede ser que además de los operadores existan otros factores de bloque que deban controlarse durante el experimento para lograr una comparación adecuada de las máquinas. También se podrían controlar: el tipo de material, lotes, tipo de producto, día, turno, etc., pero no se trata de caer en el extremo de querer controlarlo todo, sino básicamente aquellos factores que por conocimiento del proceso o experiencia previa, se sabe que afectan en forma considerable el resultado de la comparación.

Fuentes de variabilidad
Son los factores que provocan la variabilidad en los datos.

En un *diseño en bloques completos al azar* (DBCA) se consideran tres *fuentes de variabilidad*: el *factor de tratamientos*, el *factor de bloque* y el *error aleatorio*, es

decir, se tienen tres posibles "culpables" de la variabilidad presente en los datos. La palabra *completo* en el nombre del diseño se debe a que en cada bloque se prueban todos los tratamientos, o sea, los bloques están completos. La *aleatorización* se hace dentro de cada bloque; por lo tanto, no se realiza de manera total como en el diseño completamente al azar. El hecho de que existan bloques hace que no sea práctico o que incluso sea imposible aleatorizar en su totalidad.

Bloque completo
En el DBCA se refiere a que en cada bloque se prueban todos los tratamientos.

Los factores de bloqueo que aparecen en la práctica son: turno, lote, día, tipo de material, línea de producción, operador, máquina, método, etc. La imposibilidad de aleatorizar de bloque a bloque se aprecia claramente cuando se bloquean factores como *día* o *turno*, ya que no tiene sentido pensar en seleccionar al azar el orden de los días o los turnos porque es imposible regresar el tiempo.

Supongamos una situación experimental con k tratamientos y b bloques. El aspecto de los datos para este caso se muestra en la tabla 4.1, y considera una repetición en cada combinación de tratamiento y bloque.

En el ejemplo 3.1 se presenta el problema de comparar cuatro métodos de ensamble (A, B, C y D), pero si además se sospecha que los cuatro operadores que se utilizarían para realizar el ensamble pueden afectar significativamente los tiempos de ensamble y, por ende, la comparación de los métodos, entonces se debe utilizar un diseño en bloques para que la fuente adicional de variación, que representan los operadores, no vaya a sesgar las comparaciones. Esto se ve más adelante en el ejemplo 4.1.

Modelo estadístico

Cuando se decide utilizar un DBCA, el experimentador piensa que cada medición será el resultado del efecto del tratamiento donde se encuentre, del efecto del bloque al que pertenece y de cierto error que se espera sea aleatorio. El modelo estadístico para este diseño está dado por:

$$Y_{ij} = \mu + \tau_i + \gamma_j + \varepsilon_{ij}; \begin{Bmatrix} i = 1,2,\ldots,k \\ j = 1,2,\ldots,b \end{Bmatrix} \qquad \textbf{(4.1)}$$

donde Y_{ij} es la medición que corresponde al tratamiento i y al bloque j (ver tabla 4.1); μ es la media global poblacional; τ_i es el efecto debido al tratamiento i, γ_j es el efecto

Tabla 4.1 Arreglo de los datos en un diseño en bloques completos al azar.

Tratamiento	Bloque					
		1	2	3	…	*b*
	1	Y_{11}	Y_{12}	Y_{13}	…	Y_{1b}
	2	Y_{21}	Y_{22}	Y_{23}	…	Y_{2b}
	3	Y_{31}	Y_{32}	Y_{33}	…	Y_{3b}
	⋮	⋮	⋮	⋮	⋮	⋮
	k	Y_{k1}	Y_{k2}	Y_{k3}	…	Y_{kb}

debido al bloque j, y ε_{ij} es el error aleatorio atribuible a la medición Y_{ij}. Se supone que los errores se distribuyen de manera normal con media cero y varianza constante σ^2 [$N(0, \sigma^2)$], y que son independientes entre sí.

Hipótesis a probar

La hipótesis de interés es la misma para todos los diseños comparativos, y está dada por:

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k = \mu \\ H_A &: \mu_i \neq \mu_j \text{ para algún } i \neq j \end{aligned} \tag{4.2}$$

que también se puede expresar como

$$\begin{aligned} H_0 &: \tau_1 = \tau_2 = \tau_3 = \ldots = \tau_k = 0 \\ H_A &: \tau_i \neq 0 \text{ para algún } i \end{aligned} \tag{4.3}$$

En cualquiera de estas hipótesis la afirmación a probar es que la respuesta media poblacional lograda con cada tratamiento es la misma para los k tratamientos y que, por lo tanto, cada respuesta media μ_i es igual a la media global poblacional, μ. De manera alternativa, es posible afirmar que *todos los efectos de tratamiento sobre la variable de respuesta son nulos*, porque cuando el efecto $\tau_i = \mu_i - \mu = 0$, entonces necesariamente la respuesta media del tratamiento es igual a la media global ($\mu_i = \mu$).

Análisis de varianza

La hipótesis dada por (4.2 o 4.3) se prueba con un *análisis de varianza con dos criterios de clasificación*, porque se controlan dos fuentes de variación: el factor de tratamientos y el factor de bloque. En la tabla 4.2 se muestra el aspecto del ANOVA para diseño DBCA.

Los cálculos necesarios pueden ser manuales, pero siempre es más práctico hacerlos con un software estadístico, porque además proporciona muchas otras opciones gráficas y tabulares útiles (no sólo el ANOVA). Utilizando la notación de puntos

Tabla 4.2 ANOVA para un diseño en bloques completos al azar.

Fuente de variabilidad	Suma de cuadrados	Grado de libertad	Cuadrado medio	F_0	Valor-p
Tratamientos	SC_{TRAT}	$k-1$	CM_{TRAT}	$F_0 = \frac{CM_{TRAT}}{CM_E}$	$P(F > F_0)$
Bloques	SC_B	$b-1$	CM_B	$F_0 = \frac{CM_B}{CM_E}$	$P(F > F_0)$
Error	SC_E	$(k-1)(b-1)$	CM_E		
Total	SC_T	$N-1$			

vista al inicio del capítulo 3, las fórmulas más prácticas para calcular las sumas de cuadrados son:

$$SC_T = \sum_{j=1}^{b}\sum_{i=1}^{k} Y_{ij}^2 - \frac{Y_{\bullet\bullet}^2}{N}$$
$$SC_{TRAT} = \sum_{i=1}^{k} \frac{Y_{i\bullet}^2}{b} - \frac{Y_{\bullet\bullet}^2}{N} \qquad \textbf{(4.4)}$$
$$SC_B = \sum_{j=1}^{b} \frac{Y_{\bullet j}^2}{k} - \frac{Y_{\bullet\bullet}^2}{N}$$

y la del error se obtiene por sustracción como:

$$SC_E = SC_T - SC_{TRAT} - SC_B$$

Ejemplo 4.1

En el ejemplo 3.1, donde se planteó la comparación de cuatro métodos de ensamble, ahora se va a controlar activamente en el experimento a los operadores que realizarán el ensamble, lo que da lugar al siguiente diseño en bloques completos al azar.

Método	Operador			
	1	2	3	4
A	6	9	7	8
B	7	10	11	8
C	10	16	11	14
D	10	13	11	9

Recordemos que la variable de respuesta son los minutos en que se realiza el ensamble. Para comparar los cuatro métodos se plantea la hipótesis:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu$$
$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j = A, B, C, D$$

la cual se prueba mediante el análisis de varianza dado en la tabla 4.3. De esta tabla se observa que para los métodos se obtuvo un valor-p = 0.003 < α = 0.05, por lo que se rechaza la hipótesis H_0 de que el tiempo medio poblacional de los métodos de ensamble son iguales, y se acepta que al menos dos de los métodos son diferentes en cuanto al tiempo promedio que requieren. De la misma manera para operadores, como su valor-p = 0.030 < α = 0.05, el factor de bloques (operadores) también afecta, es decir, existen diferencias entre los operadores en cuanto al tiempo promedio.

Sin embargo, recordemos que no es objetivo del experimento comparar a los operadores, y su control en el estudio se utiliza para lograr una comparación más justa y precisa de los métodos de ensamble. En otras palabras, mientras que los métodos de ensamble se comparan con el objetivo final de elegir el más eficiente en términos de tiempo, con los operadores no se trata de elegir uno; en todo caso, quizá como información extra se pueda *tomar* alguna decisión sobre los operadores, como por ejemplo dar mayor entrenamiento a quien lo requiera por salirse en forma significativa del comportamiento del resto.

Cuando mediante un diseño de bloques se concluye que los tratamientos son diferentes, es probable que no se haya llegado a esa conclusión, si no que se haya considerado el factor de bloque. Por ejemplo, si en el ANOVA de la tabla 4.3 no se considera el efecto de bloque (operador), entonces la variabilidad y los grados de libertad atribuibles a operadores se irían al error, lo cual puede modificar las conclusiones sobre los tratamientos (métodos). Los detalles de esto se presentan al lector como ejercicios.

Aunque el objetivo no es que el experimentador haga los cálculos a mano, en caso de no contar con un software es posible hacer las cuentas con las fórmulas de las sumas de cuadrados dadas por la ecuación 4.4. Para calcular estas sumas es necesario obtener antes la media global y los totales por tratamiento y por bloque, como se ilustra a continuación.

	Operador				
Método	**1**	**2**	**3**	**4**	**Total por tratamiento**
A	6	9	7	8	$Y_{1\bullet} = 30$
B	7	10	11	8	$Y_{2\bullet} = 36$
C	10	16	11	14	$Y_{3\bullet} = 51$
D	10	13	11	9	$Y_{4\bullet} = 43$
Total	$Y_{\bullet 1} = 33$	$Y_{\bullet 2} = 48$	$Y_{\bullet 3} = 40$	$Y_{\bullet 4} = 39$	**Total global** $Y_{\bullet\bullet} = 160$

Con estos totales las sumas de cuadrados se obtienen fácilmente como:

$$SC_T = \sum_{j=1}^{b}\sum_{i=1}^{k} Y_{ij}^2 - \frac{Y_{\bullet\bullet}^2}{N} = \left(6^2 + 7^2 + \ldots + 9^2\right) - \frac{160^2}{16} = 108$$

$$SC_{TRAT} = \sum_{i=1}^{k} \frac{Y_{i\bullet}^2}{b} - \frac{Y_{\bullet\bullet}^2}{N} = \frac{30^2 + 36^2 + 51^2 + 43^2}{4} - \frac{160^2}{16} = 61.5 \qquad \textbf{(4.5)}$$

Tabla 4.3 ANOVA para el ejemplo 4.1.

Fuente de variabilidad	Suma de cuadrados	Grado de libertad	Cuadrado medio	F_0	Valor-*p*
Métodos	61.5	3	20.5	10.25	0.003
Operadores	28.5	3	9.5	4.75	0.030
Error	18.0	9	2.0		
Total	108.0	15			

$$SC_B = \sum_{j=1}^{b} \frac{Y_{\cdot j}^2}{k} - \frac{Y_{\cdot\cdot}^2}{N} = \frac{33^2 + 48^2 + 40^2 + 39^2}{4} - \frac{160^2}{16} = 28.5$$
$$SC_E = SC_T - SC_{TRAT} - SC_B = 18 \qquad \textbf{(4.6)}$$

Los grados de libertad de la SC_T corresponden al número total de observaciones menos uno ($N - 1 = 16 - 1 = 15$), mientras que los de las SC_{TRAT} y SC_B son el número de tratamientos menos uno y el número de operadores menos uno, respectivamente. En este caso ambas sumas tienen $4 - 1 = 3$ grados de libertad. Por último, la SC_E tiene $15 - 3 - 3 = 9$ grados de libertad. Con esta información se procede a llenar la tabla de ANOVA de la tabla 4.3.

Comparación de parejas de medias de tratamiento en el DBCA. Cuando se rechaza la hipótesis de igualdad de los cuatro tratamientos, es natural preguntarse cuáles de ellos son diferentes entre sí. Para averiguarlo se utiliza alguna de las pruebas que se estudiaron en la sección "Comparaciones o pruebas de rango múltiples" del capítulo anterior. Por ejemplo, recordemos que la diferencia mínima significativa (LSD) para dos tratamientos, i y l, en un DCA está dada por

$$LSD = t_{\alpha/2, N-k}\sqrt{\frac{2CM_E}{n}}$$

Entonces, en bloque esta expresión se transforma en

$$LSD = t_{\alpha/2, (k-1)(b-1)}\sqrt{\frac{2CM_E}{b}}$$

donde b es el número de bloques, que hace las veces de número de réplicas, y $(k - 1)(b - 1)$ son los grados de libertad del CM_E. De aquí que en el ejemplo que nos ocupa, como $t_{0.025, 9} = 2.26$, entonces,

$$LSD = 2.26\sqrt{2 \times 2 / 4} = 2.26$$

Al comparar esta diferencia mínima significativa con los datos del ejemplo 4.1 se obtiene la siguiente tabla:

Diferencia poblacional	Diferencia muestral	Decisión
$\mu_A - \mu_B$	$\lvert -1.5 \rvert < 2.26$	No significativa
$\mu_A - \mu_C$	$\lvert -5.25 \rvert > 2.26$	Significativa
$\mu_A - \mu_D$	$\lvert -3.25 \rvert > 2.26$	Significativa
$\mu_B - \mu_C$	$\lvert -3.75 \rvert > 2.26$	Significativa
$\mu_B - \mu_D$	$\lvert -1.75 \rvert < 2.26$	No significativa
$\mu_C - \mu_D$	$2.00 < 2.26$	No significativa

Se concluye que el tratamiento *A* es diferente de *C* y *D*, y que el tratamiento *B* es diferente de *C*. Las otras tres comparaciones (*A* con *B*, *B* con *D* y *C* con *D*) aceptan la hipótesis de igualdad. De acuerdo con esto, y dadas las respuestas medias muestrales $\bar{Y}_{A\cdot} = 7.5$, $\bar{Y}_{B\cdot} = 9.0$, $\bar{Y}_{C\cdot} = 10.75$, $\bar{Y}_{D\cdot} = 12.75$, se concluye que el método *A* es mejor (requiere menos tiempo para el ensamble) que los métodos *C* y *D*, pero el método *A* no es mejor que el *B*.

Efecto de bloque

Como ya vimos en el ejemplo anterior, la tabla de ANOVA también proporciona una prueba para el efecto de los bloques. En el segundo renglón de la tabla 4.3 se verifica la hipótesis

$$H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \ldots = \gamma_b = 0$$
$$H_A : \gamma_j \neq 0 \text{ para algún bloque } j$$

que en caso de rechazarse se acepta que el efecto de un bloque es diferente de cero. Por cierto, ésta no es una prueba *F* exacta, sino aproximada, debido a la restricción de aleatorización (sólo se aleatoriza dentro de bloque). Sin embargo, en la práctica se recomienda su interpretación porque es evidencia a favor o en contra de que valió la pena el esfuerzo de controlar el factor de bloque. Si resulta significativa implica que el factor de bloques tiene influencia sobre la variable de respuesta, y debe ser tomado en cuenta para mejorar la calidad de ésta. Pero, si no se rechaza y se acepta que los bloques son iguales en respuesta media, entonces se tiene el argumento a favor de no controlar este factor en futuros experimentos sobre esta misma respuesta, además de que su influencia en la calidad de la respuesta no es significativa. Por ejemplo, en este caso los operadores sí tienen efecto sobre el tiempo de ensamble, dado el valor-$p = 0.030$ que resulta en el ANOVA o, dicho en otras palabras, el tiempo medio que tardan en el ensamble los operadores es significativamente diferente. Si se hacen las comparaciones dos a dos con la prueba LSD, se encuentra que el operador 1 es estadísticamente diferente al operador 2, los demás son iguales.

La restricción de aleatorización se debe al hecho de que no se aleatoriza el orden de las corridas experimentales en relación a los bloques. El experimento supone que sólo se aleatoriza el orden de las corridas dentro de cada bloque, lo cual evita sesgos en la comparación de los tratamientos, pero no los impide en la comparación de los bloques. De hecho, todas las corridas de un bloque particular se pueden hacer de manera consecutiva, lo que puede causar sesgos a la hora de comparar los bloques. Estos sesgos se deben a factores de ruido que actúan en el transcurso de las corridas experimentales, como las variables ambientales. El error de restricción no es estimable porque se confunde con el efecto de los bloques. Por lo general se apuesta a que dicho error sea pequeño, de aquí que se recomiende interpretar la prueba *F* para los bloques dada en el ANOVA.

Si fuera de interés el estudio del factor de bloque al mismo nivel del factor de tratamientos, entonces se debería correr el experimento aleatorizando completamente el orden de todas las combinaciones posibles entre bloques y tratamientos. Si ése fuera el caso y suponiendo que sea posible aleatorizar por completo, el resultado sería un diseño factorial $k \times b$, que se presenta en el capítulo 5. Otro supuesto del

Tabla 4.4 Aspectos de los datos en un diseño en cuadro latino.

		Bloque II (columnas)				
		1	**2**	**3**	…	***k***
Bloque I (renglones)	1	$A = Y_{111}$	$B = Y_{212}$	$C = Y_{313}$	…	$K = Y_{k1k}$
	2	$B = Y_{221}$	$C = Y_{322}$	$D = Y_{423}$	…	$A = Y_{12k}$
	3	$C = Y_{331}$	$D = Y_{432}$	$E = Y_{533}$	…	$B = Y_{23k}$
	⋮	⋮	⋮	⋮	⋱	⋮
	k	$K = Y_{kk1}$	$A = Y_{1k2}$	$B = Y_{2k3}$	…	$J = Y_{jkk}$

diseño de bloques al azar es que no existe *efecto de interacción*[1] entre el factor de bloque y el factor de tratamientos. Cuando este supuesto no se cumple, la variabilidad debida a la interacción se incorpora como parte del error que, al ser grande y artificial, enmascara el efecto de los tratamientos. La existencia del efecto de interacción se puede evaluar obteniendo una suma de cuadrados aproximada para dicho efecto en el ANOVA (véase capítulo siguiente).

Efecto de interacción
Es cuando dos factores interactúan, es decir, cuando el efecto de uno depende del nivel del otro.

Diseño en cuadro latino

En el *diseño en cuadro latino* (DCL) se controlan dos factores de bloque y se estudia un factor de tratamientos, por lo que se tienen cuatro fuentes de variabilidad que pueden afectar la respuesta observada, éstas son: los tratamientos, el factor de bloque I (columnas), el factor de bloque II (renglones) y el error aleatorio. Se llama *cuadro latino* por dos razones: es un *cuadro* debido a que tiene la restricción adicional de que los tres factores involucrados se prueban en la misma cantidad de niveles, y es *latino* porque se utilizan letras latinas para denotar a los tratamientos o niveles del factor de interés. Sean A, B, C, …, K, los k tratamientos a comparar, por lo tanto ambos factores de bloques tienen también k *nivel*es cada uno. El aspecto de los datos se muestra en la tabla 4.4.

Cuadro latino
Diseño en el que se controlan dos factores de bloque y uno de tratamientos; los tres factores tienen la misma cantidad de niveles. Los tratamientos se representan por letras latinas y se distribuyen en forma adecuada en un cuadro.

Ahora se necesitan al menos tres subíndices, por ejemplo, la respuesta Y_{313} se generó en el tratamiento tres (C), en el primer nivel del factor renglón y en el tercer nivel del factor columna.

Análisis del diseño

El modelo estadístico para describir el comportamiento de las observaciones está dado por

[1] Dos factores interaccionan cuando el efecto de uno de ellos depende del nivel en que se encuentra el otro. Por ejemplo, los métodos y los operadores interactúan si la eficacia de un método depende de cuál operador lo aplique. Esto complicaría la interpretación del resultado de la comparación de los métodos, ya que debido a la interacción puede ocurrir que el mejor método no lo sea con todos los operadores, lo cual implica que para hablar del mejor método se debe decir primero de cuál operador se habla. En el siguiente capítulo se verá con detalle el concepto de *interacción*.

Tabla 4.5 ANOVA para el diseño de cuadro latino.

Fuente de variabilidad	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor-p
Tratamientos	SC_{TRAT}	$k-1$	CM_{TRAT}	$F_0 = \frac{CM_{TRAT}}{CM_E}$	$P(F > F_0)$
Renglones	SC_{B1}	$k-1$	CM_{B1}	$F_0 = \frac{CM_B}{CM_E}$	$P(F > F_0)$
Columnas	SC_{B2}	$k-1$	CM_{B2}	$F_0 = \frac{CM_{B2}}{CM_E}$	$P(F > F_0)$
Error	SC_E	$(k-2)(k-1)$	CM_E		
Total	SC_T	k^2-1			

$$Y_{ijl} = \mu + \tau_i + \gamma_j + \delta_l + \varepsilon_{ijl}$$

donde Y_{ijl} es la observación del tratamiento i, en el nivel j del factor renglón y en el *nivel l* del factor columna; ε_{ijl} es el error atribuible a dicha observación. De acuerdo con este modelo, la variabilidad total presente en los datos se puede descomponer como

$$SC_T = SC_{TRAT} + SC_{B1} + SC_{B2} + SC_E$$

y los grados de libertad correspondientes son

$$k^2 - 1 = (k-1) + (k-1) + (k-1) + (k-2)(k-1)$$

El ANOVA para el diseño en cuadro latino se muestra en la tabla 4.5. En él se prueba la hipótesis sobre los efectos de tratamiento del factor renglón y del factor columna. Otra vez, la hipótesis fundamental es la de los tratamientos; las otras dos proporcionan un adicional al objetivo inicial y permiten comprobar la relevancia de controlar los factores de bloque.

Ejemplo 4.2

Comparación de cuatro marcas de llantas. Una compañía de mensajería está interesada en determinar cuál marca de llantas tiene mayor duración en términos del desgaste. Para ello se planea un experimento en cuadro latino, en el que se comparan las cuatro marcas de llantas sometiéndolas a una prueba de 32 000 kilómetros de recorrido, utilizando cuatro diferentes tipos de auto y las cuatro posiciones posibles de las llantas en el auto. Así, el factor de interés es el *tipo de llanta* o *marca*, y se controlan dos factores de bloques: el *tipo de carro* y la *posición de la llanta en el carro*. Estos factores de bloques se controlan ya que, por experiencia, se sabe que el tipo de carro y la posición de la llanta tienen efecto en el desgaste de la misma.

La elección del cuadro latino a utilizar se hace antes de obtener los datos. Para ello, a partir de un cuadro latino inicial se aleatorizan las columnas y los renglones;

Tabla 4.6 DCL en la comparación de llantas.

Posición	Carro			
	1	2	3	4
1	$C = 12$	$D = 11$	$A = 13$	$B = 8$
2	$B = 14$	$C = 12$	$D = 11$	$A = 3$
3	$A = 17$	$B = 14$	$C = 10$	$D = 9$
4	$D = 13$	$A = 14$	$B = 13$	$C = 9$



después, las diferentes marcas de llanta se asignan de manera aleatoria a las letras latinas que denotan los niveles del factor de interés (véase la siguiente subsección).

Las pruebas se hacen al mismo tiempo con choferes, a quienes se les instruye para que manejen de manera similar sobre el mismo terreno para los cuatro automóviles. Al hacer las pruebas de los cuatro autos al mismo tiempo se evita el efecto del ambiente en el desgaste; asimismo, el conductor y el tipo de terreno podrían influir, pero se considera suficiente mantenerlos lo más homogéneos posible durante el experimento. El diseño y los datos observados se muestran en la tabla 4.6. Se mide la diferencia máxima entre el grosor de la llanta nueva y el grosor de la llanta después de haber recorrido los 32 000 kilómetros. Obviamente, a mayor diferencia en grosor mayor desgaste. Las unidades de medición son milésimas de pulgada.

Análisis de varianza. El ANOVA resultante se muestra en la tabla 4.7. Se observa que existen diferencias entre las marcas de llanta y entre los tipos de carro, a un nivel de significancia de $\alpha = 0.05$. Además, no hay evidencia suficiente para concluir que la posición tiene un efecto importante, puesto que su correspondiente valor-p es mayor que 0.05.

Interpretación. Para investigar cuáles marcas de llantas son diferentes entre sí, se aplica la prueba LSD y se obtienen los resultados de la siguiente tabla:

Marca	Núm.	Media muestral	Grupos homogéneos
C	4	10.75	X
D	4	11.00	X
B	4	12.25	X
A	4	14.25	X

Tabla 4.7 ANOVA para el ejemplo 4.2.

Fuente de variabilidad	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor-p
Marca	30.6875	3	10.2292	11.42	0.0068
Posición	6.1875	3	2.0625	2.30	0.1769
Carro	38.6875	3	12.8958	14.40	0.0038
Error	5.375	6	0.895833		
Total	80.9375	15			

Las conclusiones sobre las cuatro marcas se leen en la columna de *grupos homogéneos* como sigue: marcas con signos "X" en la misma columna son iguales estadísticamente entre sí. Por ejemplo, la marca *A* no tiene X en la primera columna y es la única con X en la segunda columna, lo cual indica que es distinta al resto de las marcas. Considerando que mientras la diferencia máxima en grosor sea mayor la llanta se desgasta más, se concluye que la marca *A* sufre mayor desgaste que las otras tres, por lo que es la peor llanta. Entre las tres marcas restantes (*C*, *D* y *B*) no se encontró una diferencia significativa en cuanto al desgaste medio. Se concluye que desde el punto de vista estadístico y a la luz de los resultados experimentales, estas tres marcas de llantas pueden considerarse iguales. Esto no quiere decir que sean idénticas, sino que sus diferencias son menores, por lo que no se alcanzan a detectar en el análisis del experimento. Dicho lo anterior, y si aún se quisiera detectar esas pequeñas diferencias para decidirse por alguna llanta, entonces habría que aumentar el número de llantas probadas, para así incrementar la potencia de la prueba. Sin embargo, quizá la mejor decisión sea no probar más llantas y decidir entre las tres marcas (*C*, *D* y *B*), con base en otros criterios, como el económico por ejemplo.

En la figura 4.1*a* se presenta la gráfica de medias para los tratamientos, donde los intervalos están construidos con el método LSD. Como ya vimos en el capítulo anterior, si los intervalos de confianza se traslapan, las respuestas medias de los tratamientos son iguales estadísticamente. Observe que el intervalo correspondiente a la marca *A* no se traslapa con ningún otro, luego, su media poblacional es diferente y mayor que las otras. Las marcas *C* y *D* son las de menor desgaste muestral y, aunque estadísticamente no difieren en media poblacional de *B*, sí hay cierta evidencia

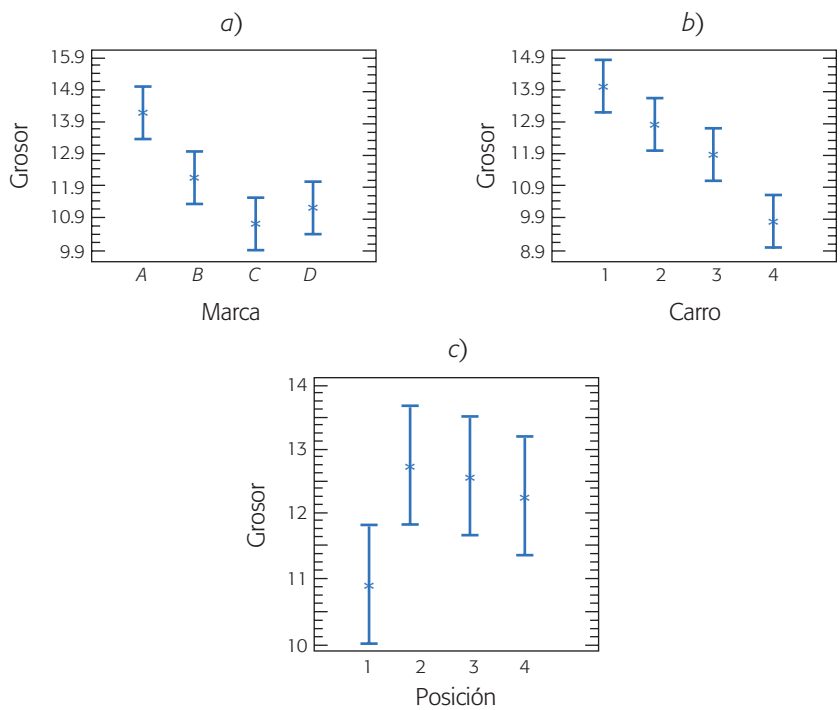


Figura 4.1 Gráficas de medias con intervalos LSD para *a*) las marcas, *b*) el carro y *c*) la posición.

(aunque no suficiente con $\alpha = 0.05$) a favor de estas dos marcas. La gráfica de medias para el factor carro (figura 4.1*b*), muestra las diferencias entre ellos: el carro 1 es el que tiene el mayor desgaste muestral de llantas, y el carro 4 es el de menor desgaste. Es posible verificar que estadísticamente son diferentes en media, el carro 4 de todos los demás y el carro 3 del carro 1.

En la figura 4.1*c* se muestra la gráfica de medias para las posiciones y todos los intervalos se traslapan, lo cual indica que no existe suficiente evidencia para concluir que las posiciones tienen algún efecto en el desgaste.

Esto es congruente con el hecho de detectar en el ANOVA que no hay efecto de la posición de las llantas. Sin embargo, se observa cierta tendencia que tiene la posición 1 a generar un menor desgaste, aunque no llega a ser significativa. Como las pruebas se hicieron en un circuito, las vueltas siempre eran en el mismo sentido, y esto puede generar mayor desgaste en una posición determinada.

Comprobación de supuestos. Como se comentó antes, la validez del análisis de varianza recae en tres supuestos que siempre deben verificarse: normalidad, varianza constante e independencia de los residuos; además de la ausencia de observaciones atípicas o aberrantes. Como se observa en la figura 4.2, el supuesto de normalidad se cumple al caer los residuos o puntos "más o menos en línea recta" (figura 4.2*a*). También se cumple el supuesto de varianza constante de acuerdo a las figuras 4.2*b* y

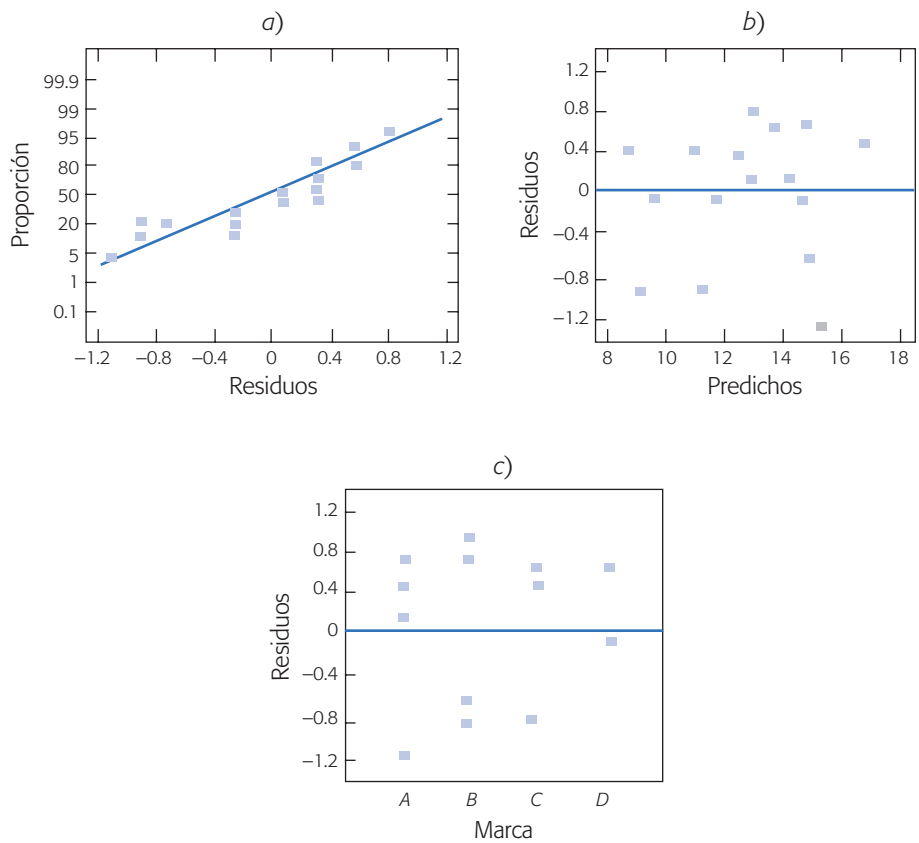


Figura 4.2 Gráficas de residuos para la verificación de supuestos del ejemplo 4.2.

4.2*c*, en las que los residuos se ubican aleatoriamente dentro de una banda horizontal; su dispersión vertical es la misma a lo largo de los gráficos. No se comprobó el supuesto de independencia porque no se conoce el orden en que se realizaron las mediciones del desgaste.

Selección y aleatorización de un cuadro latino

Cuadro latino estándar
Cuadro latino que tiene en la primera columna y en el primer renglón las letras en orden alfabético.

No cualquier arreglo de letras latinas en forma de cuadro es un cuadro latino. La regla fundamental es que cada letra debe aparecer sólo una vez en cada renglón y en cada columna. Siempre es fácil construir un *cuadro latino estándar*: en el que en la primera columna y en el primer renglón aparecen las letras en orden alfabético. Por ejemplo, un cuadro latino estándar de tamaño cuatro está dado por:

A	*B*	*C*	*D*
B	*C*	*D*	*A*
C	*D*	*A*	*B*
D	*A*	*B*	*C*

Existen además los siguientes tres cuadros latinos estándar de dimensión cuatro:

A	*B*	*C*	*D*
B	*A*	*D*	*C*
C	*D*	*B*	*A*
D	*C*	*A*	*B*

,

A	*B*	*C*	*D*
B	*D*	*A*	*C*
C	*A*	*D*	*B*
D	*C*	*B*	*A*

y

A	*B*	*C*	*D*
B	*A*	*D*	*C*
C	*D*	*A*	*B*
D	*C*	*B*	*A*

Para cuatro tratamientos se pueden construir un total de 576 cuadros latinos, de los cuales *cuatro* son estándar. La selección del diseño debería ser elegir uno al azar de los 576 posibles; no obstante, es prácticamente imposible construir todos para seleccionar uno al azar. Sin embargo, ocurre que dado un cuadro latino, cualquier intercambio de columnas o de renglones también es un cuadro latino. Por eso la estrategia de selección y aleatorización recomendada en la práctica es la siguiente:

1. Se construye el cuadro latino estándar más sencillo.
2. Se aleatoriza el orden de los renglones (o columnas) y después se aleatoriza el orden de las columnas (o renglones).
3. Por último, los tratamientos a comparar se asignan en forma aleatoria a las letras latinas.

El cuadro latino tiene dos restricciones a la aleatorización que se deben a los dos factores de bloque, lo cual implica que a la hora de correr el experimento no hay ningún margen de aleatorización. Es decir, se puede correr por columna o por renglón según convenga. Lo que no es correcto es hacer todas las pruebas de un tratamiento, luego todas las de otro, y así sucesivamente, puesto que se puede introducir ruido adicional debido a factores no controlables que cambian con el tiempo.

Diseño en cuadro grecolatino

Cuadro grecolatino
Diseño en el que se controlan tres factores de bloques y un factor de tratamiento; los cuatro factores utilizan la misma cantidad de niveles.

Con el *diseño en cuadro grecolatino* (DCGL) se controlan tres factores de bloque, además del factor de tratamientos. Se llama *cuadro grecolatino* porque los cuatro factores involucrados se prueban en la misma cantidad de niveles, de aquí que se pueda escribir como un cuadro (véase tabla 4.8); además, se utilizan letras latinas para denotar a los tratamientos y letras griegas para nombrar a los niveles del tercer factor de bloque. Al igual que en el cuadro latino, cada letra (latinas y griegas) debe aparecer sólo una vez en cada renglón y en cada columna. Además, cada par de letras debe aparecer sólo una vez en todo el arreglo. En la tabla 4.8 se presenta el aspecto de los datos del diseño en cuadro latino de dimensión $k = 4$.

El modelo estadístico que describe a las mediciones en un cuadro grecolatino está dado por

$$Y_{ijlm} = \mu + \tau_i + \gamma_j + \delta_l + \varphi_m + \varepsilon_{ijlm}$$

donde Y_{ijlm} es la observación o respuesta que se encuentra en el tratamiento i (i-ésima letra latina), en el renglón j, en la columna l y en la m-ésima letra griega; τ_i es el efecto del tratamiento i, γ_j es el efecto del renglón j, δ_l representa el efecto de la columna l y φ_m representa el efecto de la m-ésima letra griega, que son los niveles del tercer factor de bloque; el término ε_{ijlm} representa el error aleatorio atribuible a la medición Y_{ijlm}. Es importante no confundir las letras griegas del modelo que representan *efectos*, con las letras griegas en el diseño que simbolizan a los niveles del tercer factor de bloque. La variabilidad total presente en los datos se puede partir de la manera usual como

$$SC_T = SC_{TRAT} + SC_{B1} + SC_{B2} + SC_{B3} + SC_E$$

donde las sumas SC_{B1}, SC_{B2} y SC_{B3} miden la variabilidad debida a los factores de bloque renglón, columna y de letras griegas, respectivamente. Para k tratamientos, los grados de libertad correspondientes a cada suma son

$$k^2 - 1 = (k-1) + (k-1) + (k-1) + (k-3)(k-1)$$

Un bosquejo del análisis de varianza se muestra en la tabla 4.9, en la cual se prueban las hipótesis de igualdad de letras latinas (tratamientos), de renglones, de columnas y de letras griegas.

Tabla 4.8 Diseño en cuadro grecolatino.

		Columnas			
		1	2	3	4
Renglones	1	$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
	2	$B\delta$	$A\gamma$	$D\beta$	$C\alpha$
	3	$C\beta$	$D\alpha$	$A\delta$	$B\gamma$
	4	$D\gamma$	$C\delta$	$B\alpha$	$A\beta$

Tabla 4.9 ANOVA para el diseño en cuadro grecolatino.

Fuente de variabilidad	Suma de cuadrados	Grados de libertad
Tratamientos (letras latinas)	$SC_{TRAT} = \sum_{i=1}^{k} \frac{Y_{i\cdots}^2}{k} - \frac{Y_{\cdots\cdots}^2}{N}$	$k-1$
Factor de bloque I (renglones)	$SC_{B1} = \sum_{j=1}^{k} \frac{Y_{\cdot j\cdots}^2}{k} - \frac{Y_{\cdots\cdots}^2}{N}$	$k-1$
Factor de bloque II (columnas)	$SC_{B2} = \sum_{l=1}^{k} \frac{Y_{\cdots l\cdot}^2}{k} - \frac{Y_{\cdots\cdots}^2}{N}$	$k-1$
Factor de bloque III (letras griegas)	$SC_{B3} = \sum_{m=1}^{k} \frac{Y_{\cdots m}^2}{k} - \frac{Y_{\cdots\cdots}^2}{N}$	$k-1$
Error	$SC_E = SC_T - SC_{TRAT} - SC_{B1} - SC_{B2} - SC_{B3}$	$(k-3)(k-1)$
Total	$SC_T = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{l=1}^{k}\sum_{m=1}^{k} Y_{ijlm}^2 - \frac{Y_{\cdots\cdots}^2}{N}$	k^2-1

Ejemplo 4.3

En el caso de ejemplo 4.1, donde se comparan cuatro métodos de ensamble y se tiene el factor de bloque operador, se podrían tener dos factores de bloque adicionales: orden en el se hace el ensamble y lugar donde se hace. De acuerdo con esto, el diseño en cuadro grecolatino se observa en la tabla 4.10.

El análisis de varianza para el ejemplo se muestra en la tabla 4.11, en donde se aprecia que el único efecto significativo son los tratamientos (métodos), y ninguno de los factores de bloque tiene un efecto significativo sobre el tiempo de ensamble. El factor operador tiene un valor-*p* bajo, lo cual indica que podría tener un efecto significativo; sin embargo, en este experimento fue imposible detectarlo. La comparación de medias para métodos de ensamble se muestra en la siguiente tabla:

Prueba LSD para método al 95% de confianza			
Método	n_i	Media	Grupos homogéneos
A	4	7.0	X
B	4	9.25	X
C	4	12.0	X
D	4	12.75	X

donde se aprecia que los métodos *A* y *B* no son diferentes, pero sí son distintos de los métodos *C* y *D*.

Uso de software

Casi cualquier software estadístico incluye procedimientos para realizar análisis de varianza con dos criterios de clasificación.

Tabla 4.10 Diseño en cuadro grecolatino para ensamble.

		Operador			
		1	2	3	4
Orden del ensamble	1	$C\beta = 10$	$B\gamma = 10$	$D\delta = 12$	$A\alpha = 7$
	2	$B\alpha = 8$	$C\delta = 15$	$A\gamma = 7$	$D\beta = 14$
	3	$A\delta = 6$	$D\alpha = 14$	$B\beta = 11$	$C\gamma = 13$
	4	$d\gamma = 11$	$A\beta = 8$	$C\alpha = 10$	$B\delta = 8$

Los métodos son las letras latinas A, B, C y D.
El lugar de trabajo son las letras griegas α, β, γ y δ.

Tabla 4.11 ANOVA para diseño en el cuadro grecolatino de la tabla 4.10.

Fuente	Suma de cuadrados	*Gl*	Cuadrado medio	Razón *F*	Valor-*p*
Método	83.5	3	27.8333	23.86	0.0135
Operador	18.5	3	6.16667	5.29	0.1024
Orden	9.5	3	3.16667	2.71	0.2170
Lugar	2.0	3	0.666667	0.57	0.6714
Residual	3.5	3	1.16667		
Total (corregido)	117.0	15			

Las combinaciones de prueba, así como la respuesta observada se capturan manualmente en el editor de datos. Se requiere una columna por cada factor controlado en el experimento, más la columna de la variable de respuesta. Por ejemplo, en el caso del DBCA se requieren tres columnas: una para el factor de tratamientos, otra para el factor de bloques y la de la respuesta, y así sucesivamente, también se agrega una columna adicional por cada factor de bloque considerado.

Se recomienda capturar los datos y combinaciones de prueba en el orden en que se hayan realizado, ya que con ello se podrá comprobar el supuesto de independencia de los residuos.

En particular, en *Statgraphics*, la secuencia para el análisis de los diseños que usan cuando menos un factor de bloques es: *Compare* → *Analysis of variance* → *Multifactor anova.*

Existe otra posibilidad más práctica que ayuda a crear el diseño y las columnas, a la cual se accesa con la siguiente secuencia de opciones: *Special* → *Experimental Design* → *Create Design* → *Single Factor Categorical.* Después se elige el número de niveles del factor de tratamientos. Luego aparece una pantalla donde se elige el tipo de diseño a ser empleado, se incluye el diseño completamente aleatorizado y los diseños de bloque vistos en este capítulo.

En Minitab se crean columnas y se registran los datos como se indicó antes, y se sigue la secuencia *Stat* → *Anova* → *Two-way*.

Uso de Excel

El ANOVA de un diseño completamente al azar o con un criterio de clasificación se realiza con la secuencia: *Herramientas* → *Análisis de datos* → *Análisis de dos factores con una sola muestra por grupo*. Si no estuviera activada la opción de *Análisis de datos*, ésta se activa con la opción *Complementos* dentro del mismo menú de *Herramientas*. Se declara el rango de los datos, que pueden estar acomodados por columnas o por renglones. La salida contiene las estadísticas básicas de cada una de las muestras y el ANOVA correspondiente.

Preguntas y ejercicios

1. ¿En qué situaciones se aplica un diseño en bloques completos al azar? ¿En qué difieren los factores de tratamientos y de bloque?
2. ¿Qué diferencia hay entre un DBCA y los diseños en cuadro latino?
3. De acuerdo con el modelo estadístico para un diseño en bloques, ¿por qué a través de este diseño se reduce el error aleatorio?
4. A continuación se muestra parte del ANOVA para un diseño en bloques, que tiene tres tratamientos y cinco bloques con una sola repetición por tratamiento-bloque.

Fuente de variación	S. de cuadrados	G. de libertad	C. medio	Razón *F*	Valor-*p*
Tratamiento	600				
Bloque	850				
Error	500				
Total		14			

   a) Agregar en esta tabla los grados de libertad, el cuadrado medio y la razón *F* para cada una de las fuentes de variación.
   b) Interprete en forma práctica, para cada caso, lo que está estimando el cuadrado medio.
   c) Escriba el modelo estadístico y las hipótesis pertinentes.
   d) Apóyese en las tablas de la distribución *F* para aceptar o rechazar las hipótesis.
   e) Con apoyo de un software obtenga el valor-*p* para cada caso. Interprete sus resultados.

5. Realice el problema anterior, pero ahora suponga que no se bloqueó. ¿Se hubiesen obtenido las mismas conclusiones? Argumente.
6. Aunque en el análisis de varianza para un diseño en bloques completos al azar también se puede probar la hipótesis sobre si hay diferencia entre los bloques, se dice que esta hipótesis se debe ver con ciertas reservas. Explique por qué.

7. Explique por qué se utiliza el adjetivo azar en el nombre del diseño en bloques completos al azar.

8. Una compañía farmacéutica realizó un experimento para estudiar los tiempos promedio (en días) necesarios para que una persona se recupere de los efectos y complicaciones que siguen a un resfriado común. En este experimento se hizo una comparación de distintas dosis diarias de vitamina C. Para hacer el experimento se contactó a un número determinado de personas, que en cuanto les daba el resfriado empezaban a recibir algún tipo de dosis. Si la edad de las personas es una posible fuente de variabilidad, explique con detalle cómo aplicaría la idea de bloqueo para controlar tal fuente de variabilidad.

9. A continuación se muestran los datos para un diseño en bloques al azar.

		Bloque				Total por tratamiento
Tratamiento		1	2	3	4	
	A	3	4	2	6	$Y_{1\bullet} =$
	B	7	9	3	10	$Y_{2\bullet} =$
	C	4	6	3	7	$Y_{3\bullet} =$
Total por bloque		$Y_{\bullet 1} =$	$Y_{\bullet 2} =$	$Y_{\bullet 3} =$	$Y_{\bullet 4} =$	Total global = $Y_{\bullet\bullet}$

   a) Complete las sumas totales que se piden en la tabla anterior.
   b) Calcule las sumas de cuadrados correspondientes: SC_{TRAT}, SC_B, SC_T y SC_E.
   c) Obtenga la tabla de análisis de varianza y anote las principales conclusiones.
   d) Obtenga la diferencia mínima significativa (LSD) para comparar tratamientos en este diseño en bloques.

10. Se hace un estudio sobre la efectividad de tres marcas de atomizador para matar moscas. Para ello, cada producto se aplica a un grupo de 100 moscas, y se cuenta el número de moscas muertas expresado en porcentajes. Se hicieron seis réplicas, pero en días diferentes; por ello, se sospecha que puede haber algún efecto importante debido a esta fuente de variación. Los datos obtenidos se muestran a continuación:

	Número de réplica (día)					
Marca de atomizador	1	2	3	4	5	6
1	72	65	67	75	62	73
2	55	59	68	70	53	50
3	64	74	61	58	51	69

   a) Suponiendo un DBCA, formule las hipótesis adecuadas y el modelo estadístico.
   b) ¿Existe diferencia entre la efectividad promedio de los atomizadores?
   c) ¿Hay algún atomizador mejor? Argumente su respuesta.
   d) ¿Hay diferencias significativas en los resultados de diferentes días en que se realizó el experimento? Argumente su respuesta.
   e) Verifique los supuestos de normalidad y de igual varianza entre las marcas.

11. En una empresa lechera se tienen varios silos para almacenar leche (cisternas de 60 000 L). Un aspecto crítico para que se conserve la leche es la temperatura de almacenamiento. Se sospecha que en algunos silos hay problemas, por ello, durante cinco días se decide registrar la temperatura a cierta hora crítica. Obviamente la temperatura de un día a otro es una fuente de variabilidad que podría impactar la variabilidad total.

	Día				
Silo	Lunes	Martes	Miércoles	Jueves	Viernes
A	4.0	4.0	5.0	0.5	3.0
B	5.0	6.0	2.0	4.0	4.0
C	4.5	4.0	3.5	2.0	3.0
D	2.5	4.0	6.5	4.5	4.0
E	4.0	4.0	3.5	2.0	4.0

a) En este problema, ¿cuál es el factor de tratamiento y cuál el factor de bloque?
b) Suponga un DBCA, formule las hipótesis adecuadas y el modelo estadístico.
c) ¿Hay diferencia entre los silos?
d) ¿La temperatura de un día a otro es diferente?
e) Revise residuos, ¿hay algún problema evidente?

12. Se diseñó un experimento para estudiar el rendimiento de cuatro detergentes. Las siguientes lecturas de "blancura" se obtuvieron con un equipo especial diseñado para 12 cargas de lavado, distribuidas en tres modelos de lavadoras:

Detergente	Lavadora 1	Lavadora 2	Lavadora 3
A	45	43	51
B	47	44	52
C	50	49	57
D	42	37	49

a) Señale el nombre del diseño experimental utilizado.
b) Formule la hipótesis que se quiere probar en este problema.
c) Realice el análisis estadístico más apropiado para estos datos y obtenga conclusiones.

13. Con respecto al problema anterior:
a) Conteste los tres incisos del problema anterior sin tomar en cuenta el efecto de las lavadoras y obtenga conclusiones.
b) ¿Hay diferencias en las conclusiones anteriores y las del problema anterior? Explique su respuesta.
c) ¿Con cuáles conclusiones se queda? Explique su respuesta.

14. Una de las variables críticas en el proceso de ensamble del brazo lector de un disco duro es el ángulo que éste forma con el cuerpo principal de la cabeza lectora. Se corre un experimento con el objetivo de comparar dos equipos que miden dicho ángulo en unidades de radianes. Se decide utilizar como factor de bloque a los operadores de los equipos. Los resultados se muestran en la siguiente tabla:

		Equipo									
		1					2				
Operador	1	1.328, 1.113,	0.985, 1.057,	1.316, 1.144,	1.553, 1.485,	1.310 1.386	1.273, 0.789,	0.985, 0.671,	1.134, 0.554,	1.412, 1.386,	0.917 1.289
	2	1.269, 1.093,	1.268, 0.984,	1.091, 1.087,	1.195, 1.482,	1.380 1.442	1.036, 0.201,	0.783, 0.900,	1.108, 0.916,	1.129, 1.434,	1.132 1.223
	3	1.440, 1.150,	1.079, 1.190,	1.389, 1.247,	1.611, 1.617,	1.445 1.574	1.454, 1.018,	1.063, 1.050,	1.219, 0.997,	1.602, 1.538,	1.583 1.478

a) Plantee el modelo y las hipótesis más adecuadas al problema.
b) ¿Existen diferencias entre los equipos? Argumente estadísticamente.
c) ¿Existen diferencias entre los operadores?
d) Dibuje los diagramas de cajas simultáneos y las gráficas de medias para ambos factores, después interprételas.
e) Verifique los supuestos de normalidad e igualdad de varianza entre tratamientos, así como la posible presencia de puntos aberrantes.

15. Un aspecto crítico para que se conserve la leche es la temperatura de almacenamiento. De manera tradicional se han usado termómetros de mercurio (Mer) para verificar que la temperatura sea la adecuada, pero ahora se han comprado termómetros electrónicos (Rtd) para facilitar el proceso de medición. Sin embargo, se duda de las mediciones de estos nuevos dispositivos. Para aclarar dudas y diagnosticar la situación, durante cinco días se toman mediciones con ambos tipos de termómetros en varios silos (a la misma hora). Los datos para cinco silos se muestran a continuación:

	Día 1		Día 2		Día 3		Día 4		Día 5	
Silo	Mer	Rtd	Mer	Rtd	Mer	Rtd	Mer	Rtd	Mer	Rtd
A	4.0	2.6	4.0	2.8	5.0	5.0	0.5	0.0	3.0	2.4
B	5.0	6.4	6.0	6.4	2.0	2.3	4.0	4.2	4.0	4.0
C	4.5	3.3	4.0	1.4	3.5	1.8	2.0	–1.9	3.0	–7.6
D	2.5	3.1	4.0	5.0	6.5	6.6	4.5	2.7	4.0	6.3
E	4.0	0.0	4.0	0.4	3.5	0.6	2.0	–4.0	4.0	–6.3

a) Observe los datos y establezca una conjetura acerca de la confiabilidad de las mediciones con Rtd (del termómetro de mercurio no hay duda).
b) Es claro que el silo se puede ver como tratamiento y día como bloque. Considere sólo los datos de Rtd y establezca el modelo estadístico. También haga el ANOVA correspondiente y obtenga conclusiones.
c) Repita el inciso anterior pero ahora para las mediciones Mer.
d) ¿Las conclusiones obtenidas en los incisos anteriores coinciden? Comente su respuesta.
e) Datos pareados. Para comparar los dos métodos de medición (Mer y Rtd) obtenga como variable de respuesta a la diferencia de temperatura que registran los métodos para cada día en cada silo. Considerando esto, establezca el modelo estadístico, haga el ANOVA correspondiente y obtenga conclusiones.

16. Se quiere estudiar el efecto de cinco diferentes catalizadores (*A*, *B*, *C*, *D* y *E*) sobre el *tiempo de reacción* de un proceso químico. Cada lote de material sólo permite cinco corridas y cada corrida requiere aproximadamente 1.5 horas, por lo que sólo se pueden realizar cinco corridas diarias. El experimentador decide correr los experimentos con un diseño en cuadro latino para controlar activamente a los *lotes* y *días.* Los datos obtenidos son:

		Día				
		1	2	3	4	5
Lote	1	$A = 8$	$B = 7$	$D = 1$	$C = 7$	$E = 3$
	2	$C = 11$	$E = 2$	$A = 7$	$D = 3$	$B = 8$
	3	$B = 4$	$A = 9$	$C = 10$	$E = 1$	$D = 5$
	4	$D = 6$	$C = 8$	$E = 6$	$B = 6$	$A = 10$
	5	$E = 4$	$D = 2$	$B = 3$	$A = 8$	$C = 8$

a) ¿Cómo se aleatorizó el experimento?
b) Anote la ecuación del modelo y las hipótesis estadísticas correspondientes.
c) ¿Existen diferencias entre los tratamientos? ¿Cuáles tratamientos son diferentes entre sí?
d) ¿Los factores de ruido, lote y día afectan el tiempo de reacción del proceso?
e) Dibuje los gráficos de medias para los tratamientos, los lotes y los días. ¿Cuál tratamiento es mejor?
f) Verifique los supuestos del modelo, considerando que los datos se obtuvieron columna por columna, día a día.

17. En el problema anterior elimine el factor de bloque días, y conteste:
a) ¿Se justifica la eliminación?
b) Sin tomar en cuenta el día, señale el nombre del diseño, el modelo y las hipótesis más adecuadas al problema.
c) Pruebe las hipótesis y obtenga conclusiones.
d) Compare el cuadro medio del error, en este caso con el del problema anterior. ¿Qué observa? ¿Cómo lo explica?
e) ¿Por qué se obtienen las mismas conclusiones en los tratamientos de este problema y del anterior?

18. Con respecto a los problemas 16 y 17, además de eliminar el factor día ahora elimine el factor lote, y conteste lo siguiente:
a) ¿Se justifica esta segunda eliminación en ambos problemas?
b) Después de la doble eliminación, señale el nombre del diseño, el modelo y la(s) hipótesis más adecuada(s) al problema.
c) Pruebe las hipótesis y obtenga conclusiones.
d) Compare el cuadro medio del error obtenido con los de los problemas 16 y 17. ¿Qué observa? ¿Cómo lo explica?
e) ¿Por qué se obtienen las mismas conclusiones en los tratamientos de este problema y de los dos anteriores?
f) ¿Cree que esta eliminación hubiese afectado si los factores de bloque hubieran sido significativos?

19. Se comprueba el peso en gramos de un material de tres proveedores: *A*, *B* y C, por tres diferentes inspectores: I, II y III, utilizando tres diferentes escalas: 1, 2 y 3. El experimento se lleva a cabo como el siguiente cuadro latino:

Inspector	Escala		
	1	2	3
I	$A = 16$	$B = 10$	$C = 11$
II	$B = 15$	$C = 9$	$A = 14$
III	$C = 13$	$A = 11$	$B = 13$

a) ¿Hay diferencias entre los proveedores?
b) ¿Hay diferencias entre los inspectores y entre las escalas?
c) Si el peso debe ser 15 g, ¿cuál proveedor es mejor?
d) Si algún factor de bloque es no significativo, elimínelo y haga el análisis adecuado.

20. Cuando se comparan varios fertilizantes o diferentes variedades de cierto cultivo, es típico que se deba considerar el gradiente de fertilidad del suelo (factor columna) o los efectos residuales de cultivos previos (factor renglón). Considerando estos factores de bloque, Gomez y Gomez (1984) plantean un experimento en cuadro latino para com-

parar, en cuanto a rendimiento en toneladas por hectárea, tres variedades de maíz híbrido (*A*, *B*, *C*) y una variedad control (*C*). Para ello, se utiliza un campo agrícola cuadrado de 16 hectáreas, dividido en parcelas de una hectárea. Los datos de rendimiento obtenidos en cada parcela se muestran a continuación:

Ren\Col	1	2	3	4
1	1.640(*B*)	1.210(*D*)	1.425(*C*)	1.345(*A*)
2	1.475(*C*)	1.185(*A*)	1.400(*D*)	1.290(*B*)
3	1.670(*A*)	0.710(*C*)	1.665(*B*)	1.180(*D*)
4	1.565(*D*)	1.290(*B*)	1.655(*A*)	0.660(*C*)

a) ¿Existen diferencias en los rendimientos de las diferentes variedades de maíz?
b) ¿Cuál de los factores de bloque tuvo efectos?
c) ¿Se habrían detectado las mismas diferencias en los tratamientos con un diseño completamente al azar?
d) ¿Y con un diseño en bloques completos al azar?

21. Se quieren comparar tres dietas (*A*, *B*, *C*) a base de proteínas de origen vegetal utilizando 18 ratas de laboratorio de una misma camada. Primero se observa por un tiempo el apetito para formar tres grupos de seis ratas, según su voracidad; y cada uno de estos grupos se clasifica a su vez en tres grupos de dos ratas, de acuerdo a su peso inicial. Se plantea un experimento donde la variable de respuesta es el peso en gramos ganado por las ratas después de cierto periodo, con los siguientes resultados:

Apetito/ peso inicial	*A*1	*A*2	*A*3
P1	67 (*C*) 72	105 (*A*) 112	95 (*B*) 86
P2	85 (*A*) 98	75 (*B*) 67	88 (*C*) 110
P3	66 (*B*) 47	68 (*C*) 91	108 (*A*) 120

a) Analice los datos. ¿Cuáles de los factores influyen en el peso ganado por las ratas?
b) ¿Cuál dieta es mejor?
c) ¿Alguno de los factores de bloque puede ser ignorado? Argumente su respuesta.
d) Si ése fuera el caso, analice de nuevo el experimento y saque conclusiones.
e) Verifique los supuestos del modelo.

22. Una compañía distribuidora ubicada en los suburbios está interesada en estudiar la diferencia en costos (tiempo y gasolina) entre las cuatro rutas (*A*, *B*, *C*, *D*) que llevan a la zona comercial, más importante para ellos, en el otro extremo de la ciudad. Deciden correr un experimento en cuadro grecolatino controlando los factores de bloque chofer, marca de vehículo (α, β, χ, δ) y día de la semana. El experimento se repite en dos semanas diferentes, en las cuales no hay días festivos ni quincenas. Los costos observados en pesos se muestran en la siguiente tabla:

Chofer/día	Lunes	Martes	Miércoles	Jueves
Carlos	825(D, α) 750	585(C, χ) 610	550(B, β) 580	580(A, δ) 650
Enrique	650(A, χ) 725	540(B, α) 560	580(C, δ) 635	850(D, β) 770
Genaro	700(C, β) 675	650(D, δ) 740	635(A, α) 540	450(B, χ) 550
Luis	475(B, δ) 480	560(A, β) 615	650(D, χ) 725	670(C, α) 730

a) Haga el análisis de varianza de este experimento.
b) Realice las pruebas de comparaciones múltiples para los factores significativos.
c) Represente los tratamientos y factores de bloque usando gráficas de medias y diagramas de dispersión.
d) ¿Cuál es la mejor ruta? ¿Cuál es la peor?
e) ¿Hay diferencias significativas entre los choferes? ¿Y entre el tipo o marca de unidad?
f) ¿Cuáles factores de bloque valió la pena tomar en cuenta en el experimento?
g) ¿Por qué se evitaron días festivos y quincenas en el experimento? ¿Cuáles otros aspectos se tenían que tomar en cuenta?
h) Verifique los supuestos del modelo.

23. Un investigador está interesado en el efecto del porcentaje de lisina y del porcentaje de proteína en la producción de vacas lecheras. Se consideran siete niveles en cada factor.

- % de lisina: 0.0 (*A*), 0.1 (*B*), 0.2 (*C*), 0.3 (*D*), 0.4 (*E*), 0.5 (*F*), 0.6 (*G*).
- % de proteína: 2 (α), 4(β), 6(χ), 8(δ), 10(ε), 12(φ), 14(γ).

Para el estudio, se seleccionan siete vacas al azar, a las cuales se les da un seguimiento de siete periodos de tres meses. Los datos en galones de leche fueron los siguientes:

Vaca/periodo	1	2	3	4	5	6	7
1	304 ($A\alpha$)	436 ($B\varepsilon$)	350 ($C\beta$)	504 ($D\varphi$)	417 ($E\chi$)	519 ($F\gamma$)	432 ($G\delta$)
2	381 ($B\beta$)	505 ($C\varphi$)	425 ($D\chi$)	564 ($E\gamma$)	494 ($F\delta$)	350 ($G\alpha$)	413 ($A\varepsilon$)
3	432 ($C\chi$)	566 ($D\gamma$)	479 ($E\delta$)	357 ($F\alpha$)	461 ($G\varepsilon$)	340 ($A\beta$)	502 ($B\varphi$)
4	442 ($D\delta$)	372 ($E\alpha$)	536 ($F\varepsilon$)	366 ($G\beta$)	495 ($A\varphi$)	425 ($B\chi$)	507 ($C\gamma$)
5	496 ($E\varepsilon$)	449 ($F\beta$)	493 ($G\varphi$)	345 ($A\chi$)	509 ($B\gamma$)	481 ($C\delta$)	380 ($D\alpha$)
6	534 ($F\varphi$)	421 ($G\chi$)	352 ($A\gamma$)	427 ($B\delta$)	346 ($C\alpha$)	478 ($D\varepsilon$)	397 ($E\beta$)
7	543 ($G\gamma$)	386 ($A\delta$)	435 ($B\alpha$)	485 ($C\varepsilon$)	406 ($D\beta$)	554 ($E\varphi$)	410 ($F\chi$)

a) Analice este experimento. ¿Qué factores tienen efecto en la producción de leche?
b) Interprete los resultados usando gráficos de medias.
c) ¿Cómo puede explicar la falta de efectos en vacas y periodo?
d) ¿Qué porcentajes de lisina y proteína dan los mejores resultados?
e) Verifique los supuestos del modelo.

Capítulo 5
Diseños factoriales

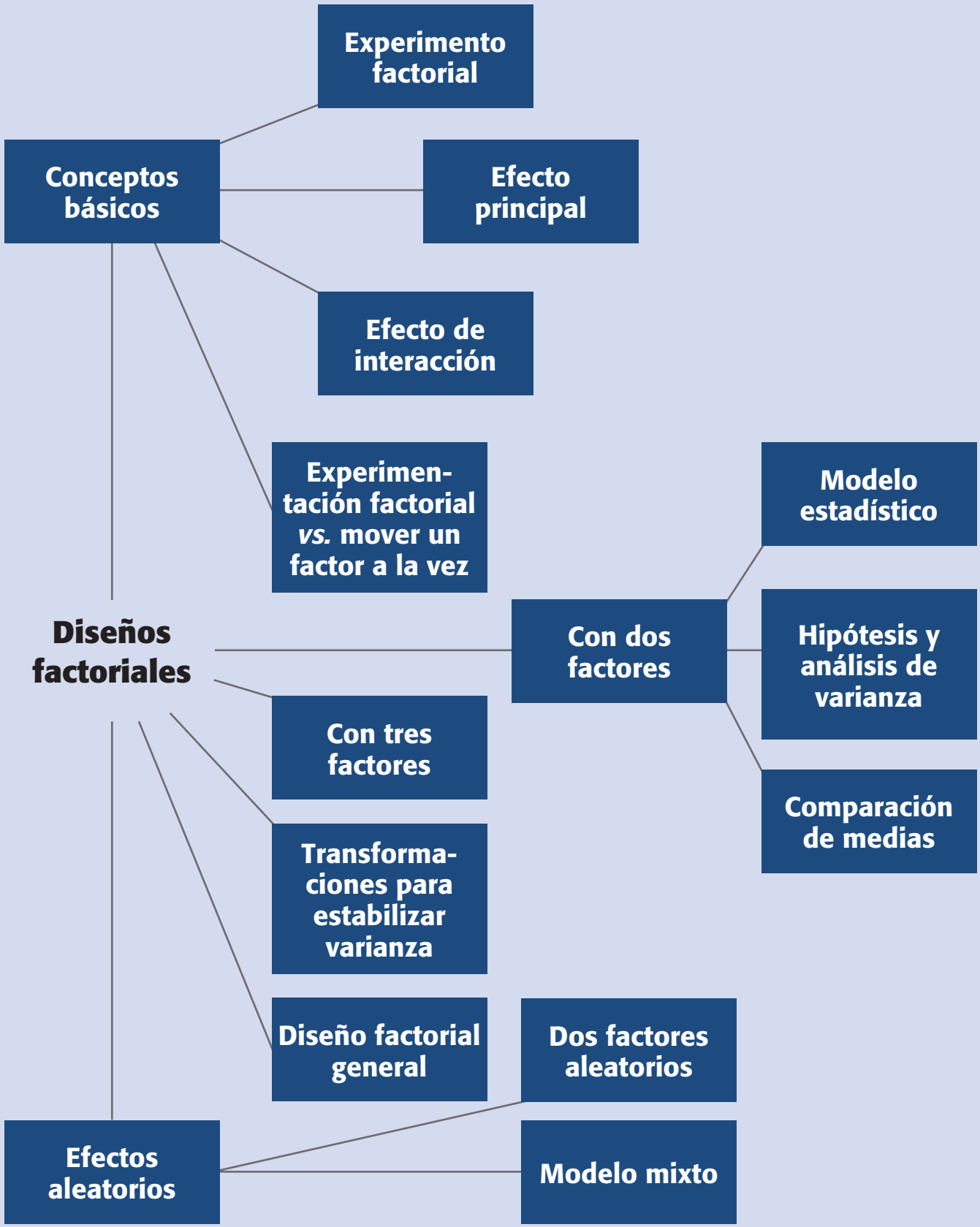
Sumario

- Conceptos básicos en diseños factoriales
- Experimentación factorial *vs.* mover un factor a la vez
- Diseños factoriales con dos factores
- Diseños factoriales con tres factores
- Transformaciones para estabilizar varianza
- Diseño factorial general
- Modelos de efectos aleatorios
- Cómo hacerlo con software

Objetivos de aprendizaje

- Describir los conceptos básicos en diseños factoriales y explicar los detalles de cómo se hace la experimentación factorial.
- Estudiar los diseños factoriales de dos y tres factores, y la manera en que se estabiliza la varianza.
- Explicar el diseño factorial general, el modelo de efectos fijos y su diferencia con el modelo de efectos aleatorios.

Mapa conceptual



Conceptos clave

- Arreglo factorial
- Comparaciones de medias
- Componentes de varianza
- Cuadrados medios
- Diseño factorial
- Efecto de interacción
- Efecto de un factor
- Efectos principales
- Factor cualitativo
- Factor cuantitativo
- Factores aleatorios
- Factores fijos
- Modelo de efectos fijos
- Principio de Pareto
- Réplica

Conceptos básicos en diseños factoriales

En el capítulo anterior se estudiaron los diseños en bloques donde sólo se tiene un factor de tratamientos, y el resto son factores de bloques que tienen una importancia secundaria en la investigación experimental. El objetivo de un *diseño factorial* es estudiar el efecto de varios factores sobre una o varias respuestas, cuando se tiene el mismo interés sobre todos los factores. Por ejemplo, uno de los objetivos particulares más importantes que en ocasiones tiene un diseño factorial es determinar una combinación de niveles de los factores en la que el desempeño del proceso sea mejor.

Los *factores* pueden ser de tipo *cualitativo* (máquinas, tipos de material, operador, la presencia o ausencia de una operación previa, etc.), o de tipo *cuantitativo* (temperatura, humedad, velocidad, presión, etc.). Para estudiar la manera en que influye cada factor sobre la variable de respuesta es necesario elegir al menos dos niveles de prueba para cada uno de ellos. Con el diseño factorial completo se corren aleatoriamente todas las posibles combinaciones que pueden formarse con los niveles de los factores a investigar.

Así, la matriz de diseño o *arreglo factorial* es el conjunto de puntos experimentales o tratamientos que pueden formarse considerando todas las posibles combinaciones de los niveles de los factores. Por ejemplo, con $k = 2$ factores, ambos con dos niveles, se forma el diseño factorial $2 \times 2 = 2^2$, que consiste en cuatro combinaciones o puntos experimentales. Si ahora uno tiene tres niveles y el otro dos, se pueden construir 3×2 combinaciones que dan lugar al diseño factorial 3×2. Observe que en el nombre del diseño factorial va implícito el número de tratamientos que lo componen. Para obtener el número de corridas experimentales se multiplica el número de tratamientos por el número de *réplicas*, donde una de éstas se lleva a cabo cada vez que se corre el arreglo completo.

En general, la familia de diseños factoriales 2^k consiste en k factores, todos con dos niveles de prueba (los factoriales 2^k se estudian con detalle en el capítulo 6); y la familia de diseños factoriales 3^k consiste en k factores cada uno con tres niveles de prueba (capítulo 7). Es claro que si los k factores no tienen la misma cantidad de niveles, debe escribirse el producto de manera explícita; por ejemplo, con $k = 3$ factores, el primero con cuatro niveles y los dos restantes con dos niveles, se tiene el diseño factorial $4 \times 2 \times 2$ o 4×2^2.

Diseño factorial
Diseño experimental que sirve para estudiar el efecto individual y de interacción de varios factores sobre una o varias respuestas.

Factor cualitativo
Sus niveles toman valores discretos o de tipo nominal. Ejemplos: máquinas, lotes, marcas, etcétera.

Factor cuantitativo
Sus niveles de prueba pueden tomar cualquier valor dentro de cierto intervalo. La escala es continua, como por ejemplo: temperatura, velocidad, presión, etcétera.

Arreglo factorial
Conjunto de puntos experimentales o tratamientos que pueden formarse al considerar todas las posibilidades de combinación de los niveles de los factores.

Ejemplo 5.1

Diseño factorial 2^2. Supongamos que en un proceso de fermentación tequilera, se tienen dos factores A: *tipo de levadura* y B: *temperatura*, cada uno con dos niveles

Tabla 5.1 Diseño factorial 2^2.

A: Levadura	*B*: Temperatura	*Y*: Rendimiento
$A_1 = 1$ (–1)	$B_1 = 22$ (–1)	28
$A_2 = 2$ (1)	$B_1 = 22$ (–1)	41
$A_1 = 1$ (–1)	$B_2 = 30$ (1)	63
$A_2 = 2$ (1)	$B_2 = 30$ (1)	45

denotados por $A_1 = 1$, $A_2 = 2$ y $B_1 = 22°C$, $B_2 = 30°C$, respectivamente. La respuesta de interés es el *rendimiento del proceso de fermentación*. En la tabla 5.1 se muestran los cuatro tratamientos o puntos del diseño factorial 2^2, y entre parentesis se ha indicado cada nivel con los códigos (1, –1). En el experimento original cada tratamiento se corrió tres veces (tres réplicas), lo cual da un total de 12 corridas del proceso pero, por simplicidad, en la última columna de la tabla 5.1 sólo se anotaron los resultados de la primera réplica.

Efecto principal y efecto de interacción. El *efecto de un factor* se define como el cambio observado en la variable de respuesta debido a un cambio de nivel de tal factor. En particular, los efectos principales son los cambios en la media de la variable de respuesta que se deben a la acción individual de cada factor. En términos matemáticos, el *efecto principal* de un factor con dos niveles es la diferencia entre la respuesta media observada cuando tal factor estuvo en su primer nivel, y la respuesta media observada cuando el factor estuvo en su segundo nivel. Por ejemplo, para los datos de la tabla 5.1, los efectos principales están dados por

$$\text{Efecto } A = \frac{41+45}{2} - \frac{28+63}{2} = -2.5$$

$$\text{Efecto } B = \frac{63+45}{2} - \frac{28+41}{2} = 19.5$$

por lo que en términos absolutos el efecto principal de B es mayor. Por otra parte, se dice que dos factores interactúan entre sí o tienen un *efecto de interacción* sobre la variable de respuesta, cuando el efecto de un factor depende del nivel en que se encuentra el otro. Por ejemplo, los factores A y B interactúan si el efecto de A es muy diferente en cada nivel de B, o viceversa. Ahora veamos esto con los datos de la tabla 5.1: el efecto de A cuando B es baja está determinado por

$$\text{Efecto } A \text{ (con } B \text{ bajo)} = 41 - 28 = 13$$

y cuando la temperatura es alta, el efecto de A es

$$\text{Efecto } A \text{ (con } B \text{ alto)} = 45 - 63 = -18$$

Como estos dos efectos de A en función del nivel de B son muy diferentes, entonces es evidencia de que la elección más conveniente del nivel de A depende del nivel en que esté B, y viceversa. Es decir, eso es evidencia de que los factores A y B interactúan sobre Y. En la práctica, el cálculo del efecto de A en cada nivel de B no se hace, y más bien se calcula el efecto global de la interacción de los dos factores, que es denotado por AB y se calculan como la diferencia entre la respuesta media cuando ambos factores se encuentran en el mismo nivel: (–1, –1); (1, 1), y la respuesta media cuando los factores se encuentran en niveles opuestos: (–1, 1) (1, –1). Para el ejemplo, el efecto de interacción *tiempo* × *temperatura* está dado por

$$AB = \frac{28+45}{2} - \frac{41+63}{2} = -15.5$$

Efecto de un factor
Es el cambio observado en la variable de respuesta debido a un cambio de nivel en el factor.

Efecto principal
Es igual a la respuesta promedio observada en el nivel alto de un factor, menos la respuesta promedio en el nivel bajo.

Efecto de interacción
Dos factores interactúan de manera significativa sobre la variable de respuesta cuando el efecto de uno depende del nivel en que está el otro.

Los valores absolutos (sin importar el signo) de los efectos principales y del efecto de interacción son una medida de importancia de su efecto sobre la variable de respuesta. Sin embargo, como se tienen estimaciones muestrales, para saber si los efectos son estadísticamente significativos (diferentes de cero) se requiere el análisis de varianza (ANOVA).

Representación de los efectos principales y la interacción

El efecto principal de un factor se representa de manera gráfica como en la figura 5.1*a*, en cuyo eje horizontal se ubican los niveles del factor y en el eje vertical se encuentra la media de la respuesta observada en los correspondientes niveles. En la figura referida se aprecia que, en el ejemplo 5.1, el efecto principal (individual) del factor *B* es mayor que el del factor *A*.

El efecto de interacción de los dos factores de la tabla 5.1 se pueden graficar como en la figura 5.1*b*; en el eje vertical se pone una escala que represente la magnitud de la variable de respuesta, luego uno de los factores se representa con sus dos niveles en el eje horizontal y en dirección vertical de cada uno de estos niveles, se anota un punto que represente la respuesta promedio en cada nivel del otro factor. Al final, cada punto del lado izquierdo se une con su correspondiente punto del lado derecho mediante una línea recta. Resulta que cuando existe interacción las líneas obtenidas tienen una pendiente muy diferente (véase figura 5.1*b*), y si no hay interacción las líneas tienen pendientes similares, que son aproximadamente paralelas (véase figura 5.2). Por ejemplo, en la figura 5.1*b* se muestra la interacción *AB*, poniendo el factor *B* en el eje horizontal, mientras que en la figura 5.1*c* se representa el mismo efecto de interacción pero ahora con el factor *A* en el eje horizontal.

Para entender e interpretar de qué manera un efecto de interacción afecta la variable de respuesta, se debe ser cuidadoso y analizar con detalle lo que pasa en *Y* cuando se mueve un factor dependiendo del nivel en el que esté el otro. Por ejemplo, en el caso del inciso *c*) de la figura 5.1 se aprecia que si *A* se cambia de su nivel (–1) al (1), cuando $B = -1$, la respuesta *Y* también se incrementa; pero si $B = 1$, la respuesta decrece de manera importante. En otras palabras, el factor *A* tiene un efecto positivo o negativo sobre *Y*, dependiendo del nivel de *B*. En el caso del inciso *b*) de esta misma figura se puede ver que si *B* se incrementa (cambia) de (–1) al (1), cuando $A = 1$, la respuesta *Y* se incrementa ligeramente; pero si $A = -1$, la respuesta se incrementa mucho.

De esta manera, con un efecto de interacción como el de la figura 5.1, si se quiere maximizar, minimizar o llevar a un valor objetivo a *Y*, no se puede mover al factor *A* sin tomar en cuenta en qué nivel está *B*, y viceversa.

Cabe señalar que en muchos procesos industriales y biológicos hay factores que tienen un efecto fuerte de interacción sobre *Y*, y que si esto se ignora, lo que algunas veces es una solución (incrementar *B* cuando $A = -1$), en otras ocasiones es insuficiente o incluso contraproducente (incrementar *B* cuando $A = 1$). Esto en la práctica provoca desconcierto a los ingenieros e investigadores, les genera explicaciones falsas y finalmente los lleva a ignorar aspectos básicos de sus procesos.

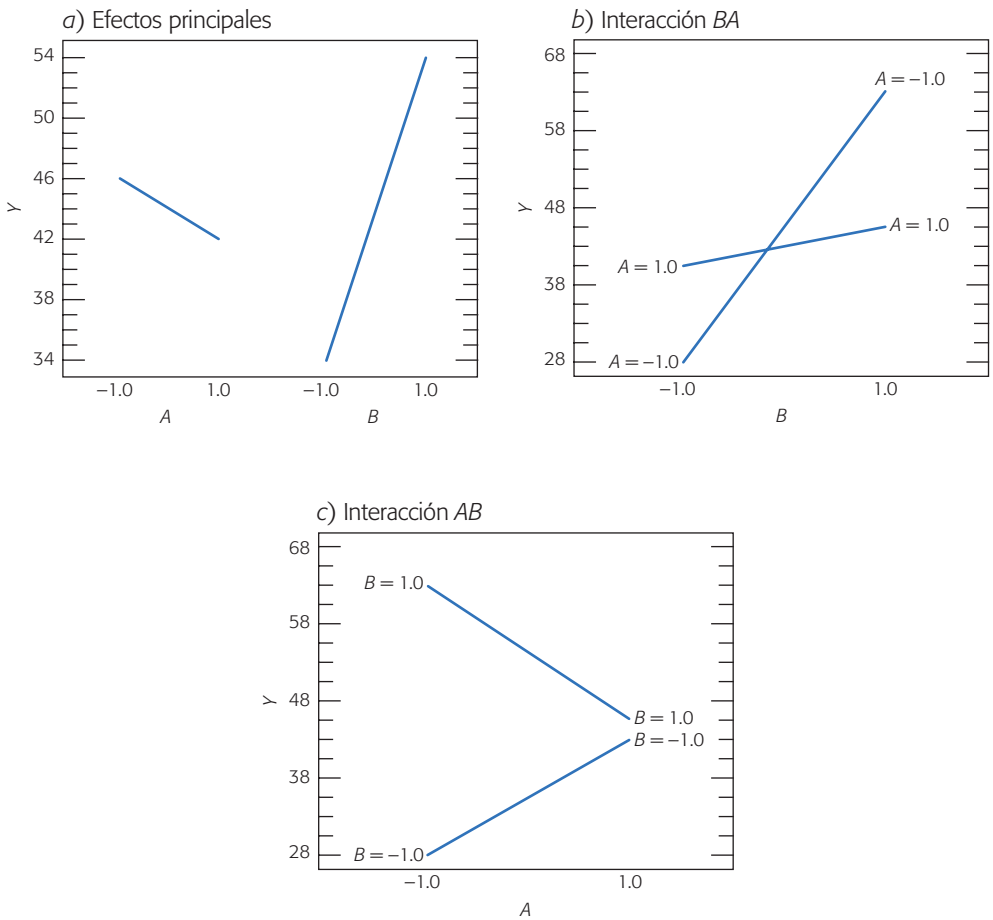


Figura 5.1 Interacción para datos tabla 5.1. Sí existe interacción: el efecto del incremento de B sobre Y es diferente dependiendo del nivel de A, y viceversa.

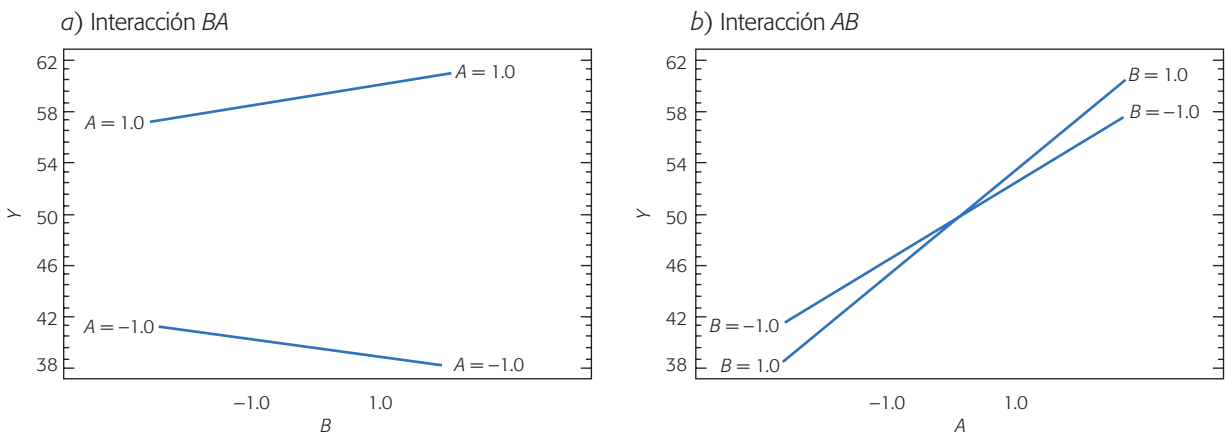


Figura 5.2 No hay efecto de interacción. En *b*) se aprecia que si *A* se aumenta, *Y* aumenta, independientemente del valor de *B*.

Ejemplo de no interacción. La interacción *AB* en la figura 5.1 contrasta con la interacción en la figura 5.2, ya que en el inciso *b*) se aprecia que si el factor *A* se cambia de su valor (–1) al (+1), la variable de respuesta *Y* aumenta de forma similar para ambos niveles del factor *B*. Mientras que en el inciso *a*) de esta figura se ve que si el factor *B* se aumenta, la respuesta *Y* prácticamente no cambia en ambos niveles de *A*. De esta manera, la figura 5.2 muestra que no hay interacción, ya que el efecto de *A* no depende del nivel donde esté *B* y viceversa.

Como hemos visto antes, la interpretación de la interacción es de vital importancia tanto para entender cómo actúan los factores sobre la variable de respuesta, como para acumular conocimiento sobre el sistema o proceso correspondiente. Cuando se concluye que una interacción doble (de dos factores) tiene un efecto estadísticamente importante sobre la respuesta, su interpretación *tiene prioridad* sobre los correspondientes efectos principales, aunque éstos también sean significativos. Esto se debe a que la interacción termina dominando en el proceso.

Una de las principales utilidades de una gráfica de interacción es que ayuda a seleccionar la condición en la que debe operarse el proceso para mejorar su desempeño. Por ejemplo, en el caso de la figura 5.1*b*, el mínimo de *Y* se logra en $A = -1$ y $B = -1$; mientras que el máximo en $A = -1$ y $B = +1$.

Experimentación factorial *vs.* mover un factor a la vez

Los diseños factoriales son más eficientes que el tradicional experimento de mover un factor a la vez, que utilizan las personas cuando no tienen conocimiento del diseño de experimentos.

Una forma de ver la ineficacia de mover un factor a la vez se ilustra a través del siguiente ejemplo. Se trata de estudiar los efectos sobre el rendimiento de un proceso que tienen tres factores: *A* (temperatura), *B* (contenido de sólidos) y *C* (tiempo de residencia). Cada factor se va a estudiar a dos niveles (–, +). Para ello, de acuerdo con el enfoque de experimentación de mover un factor a la vez, se procede de la siguiente manera:

1. Para estudiar el efecto de *A* se realizan cuatro pruebas con cada nivel de *A*, mientras que los factores *B* y *C* se fijan [en (–) por ejemplo]. Se obtiene que A_+ es mejor que A_. Véase tabla 5.2.
2. Ahora se hace lo mismo para el factor *B*, pero fijando *A* en (+) que fue el mejor y a *C* en (–). Con las cuatro pruebas en cada nivel se obtiene que con B_+ se logra un mejor rendimiento que con B_-.
3. De acuerdo con lo anterior se fija A_+ y B_+, y de igual manera se estudia el efecto de *C*. Se obtiene que C_+ es mejor que C_-. Véase tabla 5.2.
4. Conclusión: condición óptima (A_+, B_+, C_+).

Problema. La mejor condición que maximiza el rendimiento puede ser cualquiera de los cuatro tratamientos que no se probaron, como se muestra en la tabla 5.2, en la que se aprecian las ocho combinaciones o tratamientos diferentes que resultan de

tener tres factores, cada uno con dos niveles; los tratamientos que sí se probaron en cada paso y el total de corridas en las que se realizó cada tratamiento. Como se aprecia, hay tratamientos que sin saberlo fueron probados en dos pasos (8 veces), y otros en ninguno. En consecuencia, después de realizar 24 pruebas sólo se estudiaron la mitad de los posibles tratamientos. Además, la comparación entre los resultados de un tratamiento y otro no se hizo con un criterio estadístico, y no se estudió el efecto simultáneo de los factores (no se estudió la interacción). Por lo tanto, se han gastado muchos recursos y se ha obtenido poca información; asimismo, no hay garantía de que la solución propuesta sea la mejor.

El enfoque correcto. Hubiese sido mejor aplicar un diseño factorial, con el cual se investigan (en orden aleatorio) todas las posibles combinaciones de los niveles de los factores. Por ejemplo, en el caso descrito antes, con 24 pruebas o incluso con menos, 16 por decir, se podrían haber estudiado los ocho tratamientos, y después de un análisis estadístico adecuado, saber qué factores son más importantes, si interactúan sobre Y y concluir cuál es el tratamiento que más conviene.

Por lo anterior, la experimentación empírica en apariencia más simple y sencilla ha resultado cara y poco eficaz. Esto se agrava a medida que se tienen más factores.

Ventajas de los diseños factoriales

1. Permiten estudiar el efecto individual y de interacción de los distintos factores.
2. Son diseños que se pueden aumentar para formar diseños compuestos en caso de que se requiera una exploración más completa. Por ejemplo, es útil aumentar el diseño si el comportamiento de la respuesta no es lineal en los factores controlados (capítulo 12).
3. Se pueden correr fracciones de diseños factoriales, las cuales son de gran utilidad en las primeras etapas de una investigación que involucra a muchos factores, cuando interesa descartar de manera económica los que no son

Tabla 5.2 Ejemplo de experimentación empírica (requiere más recursos).

Tratamiento	*A*	*B*	*C*	Tratamientos probados			Total de corridas para cada tratamiento
				Paso 1	Paso 2	Paso 3	
1	–	–	–	*			4
2	+	–	–	*	*		8
3	–	+	–				0
4	+	+	–		*	*	8
5	–	–	+				0
6	+	–	+				0
7	–	+	+				0
8	+	+	+			*	4
						Total	24

importantes, antes de hacer un estudio más detallado con los factores que sí son importantes (capítulo 8).

4. Pueden utilizarse en combinación con diseños de bloques en situaciones en las que no puede correrse todo el diseño factorial bajo las mismas condiciones. Por ejemplo, cuando cada lote de material sólo alcanza para correr la mitad del experimento, éste se puede realizar en dos bloques (dos lotes), lo cual implica repartir las pruebas en los dos lotes de la manera más conveniente posible.
5. La interpretación y el cálculo de los efectos en los experimentos factoriales se puede hacer con aritmética elemental, en particular cuando cada factor se prueba en dos niveles.

Todas estas ventajas de los diseños factoriales se harán patentes en la medida en que nos adentremos en el tema a lo largo del libro.

Diseños factoriales con dos factores

Réplica
Es cada corrida de todos los tratamientos del arreglo factorial.

Considere los factores A y B con a y b $(a, b \geq 2)$ niveles de prueba, respectivamente. Con ellos se puede construir el arreglo o diseño factorial $a \times b$, el cual consiste en $a \times b$ tratamientos. Algunos casos particulares de uso frecuente son: el factorial 2^2, el factorial 3^2 y el factorial 3×2. Se llama *réplica* a cada corrida completa del arreglo factorial. Los diseños factoriales que involucran menos de cuatro factores por lo regular se corren replicados para tener la potencia necesaria en las pruebas estadísticas sobre los efectos de interés. Si se hacen n réplicas, el número total de corridas experimentales es $n(a \times b)$.

Ejemplo 5.2

Factorial 4 × 3. Consideremos un experimento en el que se quiere estudiar el efecto de los factores A: *profundidad de corte* sobre el *acabado* de un metal y B: *velocidad de alimentación*. Aunque los factores son de naturaleza continua, en este proceso sólo se puede trabajar en 4 y 3 niveles, respectivamente. Por ello, se decide correr un factorial completo 4×3 con tres réplicas, que permitirá obtener toda la información relevante en relación al efecto de estos factores sobre el acabado. Al aleatorizar las 36 pruebas se obtienen los datos de la tabla 5.3.

El acabado (Y) está en unidades de gramos e interesa minimizar su valor. La representación gráfica del diseño 4×3 se muestra en la figura 5.3.

Modelo estadístico e hipótesis de interés. El modelo estadístico de efectos para este tipo de diseño está dado por:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}; \quad i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, b;\ k = 1, 2, \ldots, n \tag{5.1}$$

donde μ es la media general, α_i es el efecto debido al i-ésimo nivel del factor A, β_j es el efecto del j-ésimo nivel del factor B, $(\alpha\beta)_{ij}$ representa al efecto de interacción en la combinación ij y ε_{ijk} es el error aleatorio que se supone sigue una distribución

Tabla 5.3 Datos del experimento factorial 4 × 3 (ejemplo 5.2).

		***B*: velocidad**			
		0.20	**0.25**	**0.30**	**Total $Y_{i..}$**
A: profundidad	0.15	74 64 [198] 60	92 86 [266] 88	99 98 [299] 102	763
	0.18	79 68 [220] 73	98 104 [290] 88	104 99 [298] 95	808
	0.21	82 88 [262] 92	99 108 [302] 95	108 110 [317] 99	881
	0.24	99 104 [299] 96	104 110 [313] 99	114 111 [332] 107	944
	Total $Y_{.j.}$	979	1 171	1 246	$Y_{...}$ = 3 396

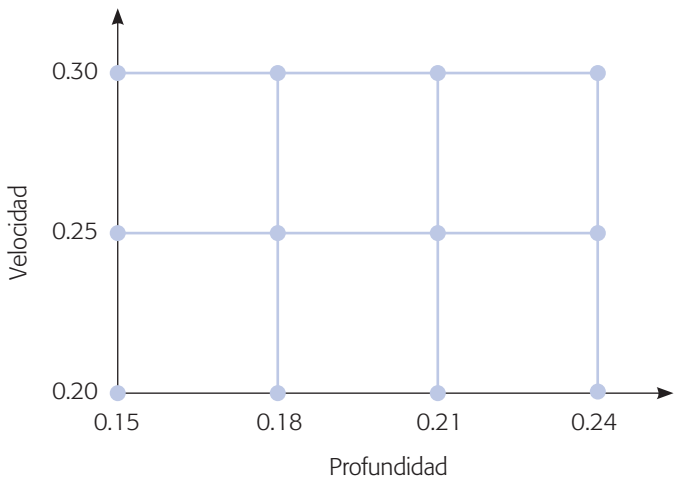


Figura 5.3 Representación del diseño factorial 4 × 3.

normal con media cero y varianza constante σ^2 ($N(0, \sigma^2)$) y son independientes entre sí. Para que la estimación de los parámetros en este modelo sea única, se introducen las restricciones $\Sigma_{i=1}^{a}\alpha_i = 0$, $\Sigma_{j=1}^{b}\beta_j = 0$ y $\Sigma_{i=1}^{a}\Sigma_{j=1}^{b}(\alpha\beta)_{ij} = 0$. Es decir, los efectos dados en el modelo son desviaciones relacionadas con la media global.

En este modelo para el ejemplo 5.2 $a = 4$, $b = 3$ y $n = 3$ replicas. Las hipótesis de interés para los tres efectos en el modelo anterior son:

$$H_0 : \text{Efecto de profundidad } (A) = 0$$
$$H_A : \text{Efecto de profundidad } (A) \neq 0$$

$$H_0 : \text{Efecto de velocidad } (B) = 0$$
$$H_A : \text{Efecto de velocidad } (B) \neq 0$$

$$H_0: \text{Profundidad} \times \text{velocidad } (AB) = 0$$
$$H_A: \text{Profundidad} \times \text{velocidad } (AB) \neq 0$$

Estas hipótesis también se pueden plantear con los efectos descritos en el modelo (5.1):

$$H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0$$
$$H_A: \alpha_i \neq 0 \text{ para algún } i$$

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_b = 0$$
$$H_A: \beta_j \neq 0 \text{ para algún } i$$

$$H_0: (\alpha\beta)_{ij} = 0 \text{ para todo } ij$$
$$H_A: (\alpha\beta)_{ij} \neq 0 \text{ para algún } ij$$

Estas hipótesis se prueban mediante la técnica de análisis de varianza, que para un diseño factorial $a \times b$ con n réplicas resulta de descomponer la variación total como,

$$SC_T = SC_A + SC_B + SC_{AB} + SC_E$$

donde los respectivos grados de libertad de cada una de ellas son:

$$nab - 1 = (a-1) + (b-1) + (a-1)(b-1) + ab(n-1)$$

El factor $(n - 1)$ en los grados de libertad de la suma de cuadrados del error (SC_ε) señala que se necesitan al menos dos réplicas del experimento para calcular este componente y, por ende, para construir una tabla de ANOVA. Recordemos que las sumas de cuadrados divididas entre sus correspondientes grados de libertad se llaman *cuadrados medios* (*CM*). Al dividir éstos entre el cuadrado medio del error (CM_E) se obtienen estadísticos de prueba con distribución F. Toda esta información se sintetiza en la tabla 5.4.

Cuadrados medios
Las sumas de cuadrados divididos entre sus correspondientes grados de libertad.

Si el valor-p es menor al nivel de significancia α prefijado, se rechaza la hipótesis nula y se concluye que el correspondiente efecto está activo o influye en la variable de respuesta.

Tabla 5.4 ANOVA para el diseño factorial $a \times b$.

FV	SC	GL	CM	F_0	Valor-p
Efecto A	SC_A	$a-1$	CM_A	CM_A/CM_E	$P(F > F_0^A)$
Efecto B	SC_B	$b-1$	CM_B	CM_B/CM_E	$P(F > F_0^B)$
Efecto AB	SC_{AB}	$(a-1)(b-1)$	CM_{AB}	CM_{AB}/CM_E	$P(F > F_0^{AB})$
Error	SC_E	$ab(n-1)$	CM_E		
Total	SC_T	$abn-1$			

Recordemos del capítulo 3 la notación de puntos para representar sumas y medias: $Y_{\cdots}$ es la suma de todas las observaciones; $\bar{Y}_{\cdots}$ es la media global; $Y_{i\cdot\cdot}$ es el total en el nivel i del factor A; $\bar{Y}_{i\cdot\cdot}$ es la media en el nivel i del factor A; $Y_{\cdot j\cdot}$ es el total en el nivel j del factor B y $\bar{Y}_{\cdot j\cdot}$ es la correspondiente media. Es decir:

$$Y_{\cdots} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} \qquad \bar{Y}_{\cdots} = \frac{Y_{\cdots}}{abn}$$

$$Y_{i\cdot\cdot} = \sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} \qquad \bar{Y}_{i\cdot\cdot} = \frac{Y_{i\cdot\cdot}}{bn} \qquad i = 1, 2, \ldots, a$$

$$Y_{\cdot j\cdot} = \sum_{i=1}^{a}\sum_{k=1}^{n} Y_{ijk} \qquad \bar{Y}_{\cdot j\cdot} = \frac{Y_{\cdot j\cdot}}{an} \qquad j = 1, 2, \ldots, b$$

$$Y_{ij\cdot} = \sum_{k=1}^{n} Y_{ijk} \qquad \bar{Y}_{ij\cdot} = \frac{Y_{ij\cdot}}{n}$$

Con esta notación la suma de cuadrados totales es:

$$SC_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk}^2 - \frac{Y_{\cdots}^2}{N} \qquad \textbf{(5.2)}$$

donde $N = abn$ es el total de observaciones en el experimento. Las sumas de cuadrados de efectos son:

$$SC_A = \sum_{i=1}^{a} \frac{Y_{i\cdot\cdot}^2}{bn} - \frac{Y_{\cdots}^2}{N},$$

$$SC_B = \sum_{j=1}^{b} \frac{Y_{\cdot j\cdot}^2}{an} - \frac{Y_{\cdots}^2}{N},$$

$$SC_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{Y_{ij\cdot}^2}{n} - \frac{Y_{\cdots}^2}{N} - SC_A - SC_B$$

y al final, al restar éstas del total, se obtiene la suma de cuadrados del error como:

$$SC_E = SC_T - SC_A - SC_B - SC_{AB} \qquad \textbf{(5.3)}$$

De acuerdo a esto para obtener el ANOVA para el ejemplo 5.2, en la tabla 5.3 se han calculado los totales necesarios. De donde:

$$SC_A = \sum_{i=1}^{4} \frac{Y_{i\cdot\cdot}^2}{3\times 3} - \frac{Y_{\cdots}^2}{4\times 3\times 3} = \frac{(763^2 + 808^2 + 881^2 + 944^2)}{3\times 3} - \frac{3396^2}{4\times 3\times 3} = 2\,125.1$$

$$SC_B = \sum_{j=1}^{3} \frac{y_{\cdot j\cdot}^2}{4\times 3} - \frac{y_{\cdots}^2}{4\times 3\times 3} = \frac{(979^2 + 1\,171^2 + 1\,246^2)}{4\times 3} - \frac{3396^2}{4\times 3\times 3} = 3\,160.5$$

$$SC_{AB} = \sum_{j=1}^{4}\sum_{i=1}^{3} \frac{y_{ij\cdot}^2}{3} - \frac{y_{\cdots}^2}{4\times 3\times 3} - SC_A - SC_B$$

$$= \frac{(198^2 + 220^2 + \ldots + 332^2)}{3} - \frac{3\,396^2}{4\times 3\times 3} - 2\,125.1 - 3\,160.5 = 557.07$$

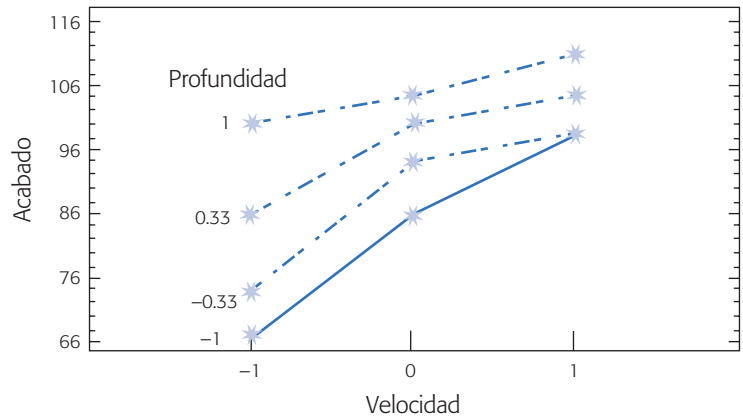


Figura 5.4 Efecto de interacción velocidad × profundidad, ejemplo 5.2.

La suma de cuadrados totales y la suma de cuadrados del error están dadas por

$$SC_T = \sum_{i=1}^{4}\sum_{j=1}^{3}\sum_{l=1}^{3} Y_{ijl}^2 - \frac{Y_{...}^2}{4\times 3\times 3} = 6\,532.0$$

$$SC_E = SC_T - SC_A - SC_B - SC_{AB} = 6\,532.0 - 2\,125.1 - 3\,160.5 - 557.07 = 689.33$$

Con esta información se construye el análisis de varianza de la tabla 5.5.

Del ANOVA se concluye que los tres efectos A : *vel*, B : *prof* y AB están activos o influyen en el acabado. Dado que el efecto de interacción AB resulta significativo, prácticamente toda la información relevante del experimento se aprecia en su representación gráfica (figura 5.4). Nótese que aparecen tantas líneas como niveles tenga el factor que se dibuja en la parte de arriba, que en este caso es la profundidad con sus cuatro niveles que se denotan con una escala de –1 a 1. La significancia de la interacción detectada por el ANOVA se observa en el hecho de que las líneas en la figura 5.4 tienen pendientes relativamente diferentes. Como lo que interesa es minimizar la variable de respuesta, se observa que a mayor velocidad y profundidad hay una tendencia a obtener peores acabados. Además se ve que cuando se tiene velocidad alta (A^+) el efecto de profundidad es menor (véase la dispersión de las líneas en la figura cuando la velocidad es alta). Por lo tanto, las condiciones de operación o tratamiento que convienen es profundidad y velocidad bajas (A^-, B^-).

El ANOVA de la tabla 5.5 se dice que no está desglosado, ya que cuando en un experimento hay factores cuantitativos con más de dos niveles, el ANOVA se puede

Tabla 5.5 ANOVA para el ejemplo 5.2.

FV	SC	GL	CM	F_0	Valor-*p*
B: veloc	3 160.5	2	1 580.25	55.02	0.0000
A: prof	2 125.10	3	708.37	24.66	0.0000
AB	557.07	6	92.84	3.23	0.0180
Error	689.33	24	28.72		
Total	6 532.0	35			

desglosar para estudiar con mayor detalle el efecto de tal factor. Esto se verá con detalle en el capítulo 7.

Comparación de medias

Comparaciones de medias
Técnicas que tienen como objeto investigar las medias que causan las diferencias detectadas en el ANOVA.

Las *comparaciones de medias* se introdujeron en la sección "Diseño completamente al azar y ANOVA" del capítulo 3, para después de un ANOVA en el que se rechaza H_0, investigar cuáles medias causan las diferencias detectadas. El ANOVA sólo indica que al menos un par de niveles del factor significativo son diferentes entre sí, pero no dice cuáles son. Por facilidad, denotemos los cuatro niveles de la profundidad (A) en el ejemplo 5.2 como A_1, A_2, A_3 y A_4, así como los tres niveles de la velocidad (B) como B_1, B_2 y B_3. Entonces, los seis pares de hipótesis para comparar las medias del factor A son:

$$\begin{array}{lll} H_0 : \mu_{A_1} = \mu_{A_2} & H_0 : \mu_{A_1} = \mu_{A_3} & H_0 : \mu_{A_1} = \mu_{A_4} \\ H_A : \mu_{A_1} \neq \mu_{A_2}\ ; & H_A : \mu_{A_1} \neq \mu_{A_3}\ ; & H_A : \mu_{A_1} \neq \mu_{A_4}\ , \\ H_0 : \mu_{A_2} = \mu_{A_3} & H_0 : \mu_{A_2} = \mu_{A_4} & H_0 : \mu_{A_3} = \mu_{A_4} \\ H_A : \mu_{A_2} \neq \mu_{A_3}\ ; & H_A : \mu_{A_2} \neq \mu_{A_4} \quad \text{y} & H_A : \mu_{A_3} \neq \mu_{A_4} \end{array}$$

mientras que para el factor B se tienen los tres pares de hipótesis,

$$\begin{array}{lll} H_0 : \mu_{B_1} = \mu_{B_2} & H_0 : \mu_{B_2} = \mu_{B_3} & H_0 : \mu_{B_1} = \mu_{B_3} \\ H_A : \mu_{B_1} \neq \mu_{B_2}\ ; & H_A : \mu_{B_2} \neq \mu_{B_3} \quad \text{y} & H_A : \mu_{B_1} \neq \mu_{B_3} \end{array}$$

Para probar estas hipótesis con el método LSD (véase capítulo 3) habría que calcular las diferencias muestrales en valor absoluto y compararlas con la diferencia mínima significativa. Cabe aclarar que este análisis es engañoso cuando el efecto de interacción es significativo, como es el caso del ejemplo 5.2. Por ello, y sólo por ilustrar el método, se prueban las hipótesis del factor A ignorando por el momento la interacción. La diferencia mínima significativa para comparar los niveles i y l del factor A, está dada por:

$$LSD_A = t_{\alpha/2,\, ab(n-1)} \sqrt{CM_E \left(\frac{1}{n_{A_i}} + \frac{1}{n_{A_l}} \right)} \qquad \textbf{(5.4)}$$

donde $t_{\alpha/2,\, ab(n-1)}$ es el punto porcentual $100(1 - \alpha/2)$ de la distribución T de Student, $ab(n - 1)$ los grados de libertad del cuadrado medio del error (tabla 5.5), n_{A_i} y n_{A_l} son el total de observaciones en los niveles i y l del factor A, que se están comparando. De esta manera, en el ejemplo, como es un diseño balanceado $n_{A_i} = n_{A_l} = 9$; entonces,

$$LSD_A = 2.064\sqrt{28.72(2/9)} = 5.21$$

De los totales marginales dados en el renglón inferior de la tabla 5.3 de datos, se obtienen las medias del factor A, al dividir entre 9, que son el número de mediciones

involucradas en cada total. Así, las seis posibles diferencias muestrales en valor absoluto resultan ser:

$$\left|\bar{Y}_{A_1} - \bar{Y}_{A_2}\right| = \frac{1}{9}|763 - 808| = 5.0 < LSD_A$$

$$\left|\bar{Y}_{A_1} - \bar{Y}_{A_3}\right| = \frac{1}{9}|763 - 881| = 13.1^* > LSD_A$$

$$\left|\bar{Y}_{A_1} - \bar{Y}_{A_4}\right| = \frac{1}{9}|763 - 944| = 20.1^* > LSD_A$$

$$\left|\bar{Y}_{A_2} - \bar{Y}_{A_3}\right| = \frac{1}{9}|808 - 881| = 8.1^* > LSD_A$$

$$\left|\bar{Y}_{A_2} - \bar{Y}_{A_4}\right| = \frac{1}{9}|808 - 944| = 15.1^* > LSD_A$$

$$\left|\bar{Y}_{A_3} - \bar{Y}_{A_4}\right| = \frac{1}{9}|881 - 944| = 7.0^* > LSD_A$$

donde sólo la primer diferencia resulta no significativa, es decir, se acepta $H_0 : \mu_{A_1} = \mu_{A_2}$; en cambio, en las cinco comparaciones restantes se rechaza H_0.

Las conclusiones, para ambos factores usando el método LSD se observan en las gráficas de medias de la figura 5.5, donde no se toma en cuenta el efecto de interacción detectado en el ANOVA.

Tomando en cuenta la interacción

Para hacer comparaciones múltiples de medias de un factor, tomando en cuenta el efecto de interacción, éstas se realizan de manera separada en cada nivel del otro factor. Por ejemplo, las comparaciones que acabamos de hacer para el factor *A* se realizan dentro de cada nivel del factor *B*; de esta forma, se toma en cuenta el efecto de interacción, y por ende, se tiene una interpretación más cercana a la realidad del proceso.

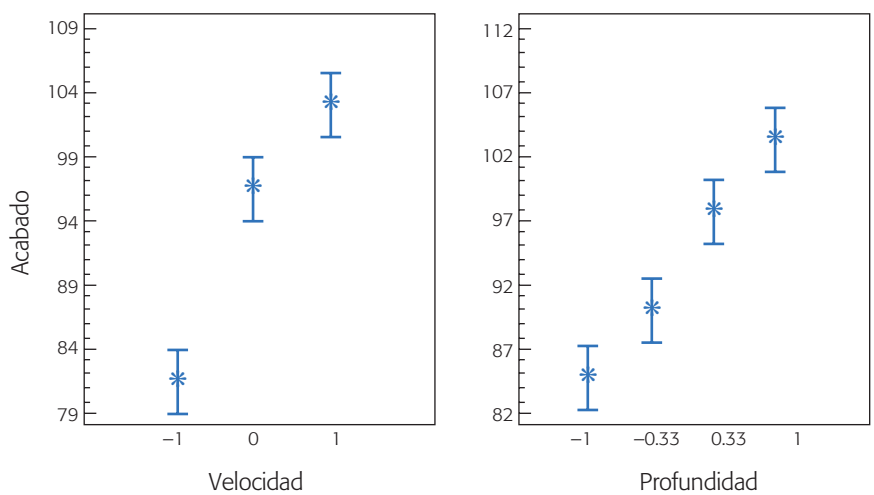


Figura 5.5 Gráficas de medias para velocidad y profundidad; ejemplo 5.2.

Si observamos el efecto de interacción en la figura 5.4, es fácil notar que las medias de las tres últimas profundidades (denotadas –0.33, 0.33 y 1) están más cercanas entre sí cuando la velocidad está en su nivel intermedio que cuando está en su nivel bajo. Veámoslo de manera analítica en la velocidad intermedia (B_2) (tabla 5.3), donde las medias muestrales del factor A : *prof* en la velocidad intermedia son:

$$\bar{Y}_{1,2,\bullet} = \frac{266}{3} = 88.66;\ \bar{Y}_{2,2,\bullet} = \frac{290}{3} = 96.66;\ \bar{Y}_{3,2,\bullet} = \frac{302}{3} = 100.66 \text{ y } \bar{Y}_{4,2,\bullet} = \frac{313}{3} = 104.33$$

Entonces, para comparar estas medias la diferencia mínima significativa está dada por

$$LSD_{B_2(A)} = t_{\alpha/2,\,ab(n-1)}\sqrt{CM_E\left(\frac{1}{n}+\frac{1}{n}\right)}$$

donde n es el número de réplicas de los tratamientos a comparar. Note que la diferencia entre esta expresión para la LSD y la que no toma en cuenta la interacción, expresión (5.4), está precisamente en el factor dentro de la raíz cuadrada que acompaña al CM_E, ya que ahí se anota el inverso del número de observaciones con las que se construyeron las medias muestrales para calcular las diferencias. Así en este caso $n = 3$, y:

$$LSD_{B_2(A)} = 2.064 \times \sqrt{28.72 \times \left(\frac{2}{3}\right)} = 9.03$$

Al comparar las seis posibles diferencias de medias de los niveles de A en valor absoluto contra la cota $LSD_{B_2(A)}$ resulta lo siguiente:

$$\left|\bar{Y}_{B_2(A_1)} - \bar{Y}_{B_2(A_2)}\right| = |88.66 - 96.66| = 8.0 < LSD_{B_2(A)}$$
$$\left|\bar{Y}_{B_2(A_1)} - \bar{Y}_{B_2(A_3)}\right| = |88.66 - 100.66| = 12.0^* > LSD_{B_2(A)}$$
$$\left|\bar{Y}_{B_2(A_1)} - \bar{Y}_{B_2(A_4)}\right| = |88.66 - 104.33| = 15.7^* > LSD_{B_2(A)}$$
$$\left|\bar{Y}_{B_2(A_2)} - \bar{Y}_{B_2(A_3)}\right| = |96.66 - 100.66| = 4.0 < LSD_{B_2(A)}$$
$$\left|\bar{Y}_{B_2(A_2)} - \bar{Y}_{B_2(A_4)}\right| = |96.66 - 104.33| = 7.7 < LSD_{B_2(A)}$$
$$\left|\bar{Y}_{B_2(A_3)} - \bar{Y}_{B_2(A_4)}\right| = |100.66 - 104.33| = 3.7 < LSD_{B_2(A)}$$

Por lo tanto, al tomar en cuenta el efecto de interacción AB se concluye que cuando $B = B_2$, sólo hay diferencias entre el nivel A, con A_3 y A_4. La gráfica de interacción con los intervalos LSD sobrepuestos se muestra en la figura 5.6. Note que en el nivel intermedio de la velocidad, los intervalos de confianza para las medias de las tres profundidades superiores se traslapan, lo cual es un indicio de que son estadísticamente iguales, como se acaba de concluir de manera analítica.

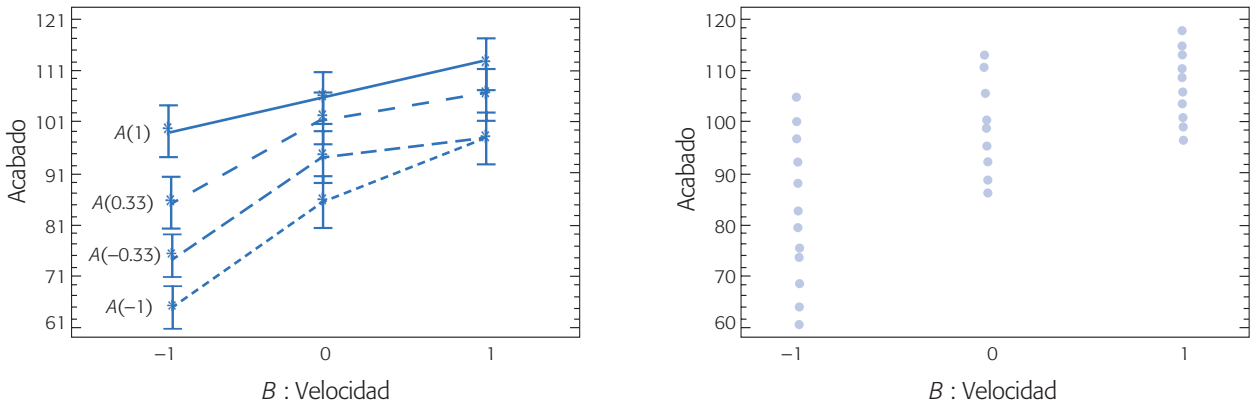


Figura 5.6 Gráfica de interacción *A* : *prof* × *B* : *vel* con intervalos *LSD* sobrepuestos y diagrama de dispersión (*B*, *Y*).

Otro aspecto importante cuando los factores son cuantitativos y se pueden apreciar en las diferentes gráficas: de interacción, de medias y de dispersión (figuras 5.5 y 5.6), es la posible presencia de curvatura en el factor velocidad. Note que en la gráfica de medias para velocidad (figura 5.5), no es posible trazar una recta que pase por el interior de los tres intervalos de confianza, lo cual es evidencia a favor de que existe curvatura en el factor *B* : *vel*. En cambio, sí es posible trazar dicha recta a través de los intervalos para el factor *A* : *prof*, lo cual es evidencia en contra de la presencia de curvatura en este factor. Esta posible curvatura se evalúa analíticamente con el ANOVA desglosado que se presenta para este mismo ejemplo en el capítulo 7.

Cabe señalar que, en general, las representaciones gráficas de efectos principales y de interacción significativos son suficientes para elegir el mejor tratamiento del experimento. La gráfica de dispersión (*Velocidad*, *Acabado*) presentada en la figura 5.6 es útil porque permite observar la variabilidad de la respuesta en cada punto experimental, además detecta la posible presencia de observaciones atípicas que pueden afectar los resultados y, por ende, las conclusiones del estudio. Todavía falta verificar los supuestos del modelo y hacer corridas confirmatorias del proceso sobre el mejor tratamiento, antes de proponer a éste como la forma estándar de operar el proceso.

Verificación de supuestos

Los *supuestos* de normalidad, varianza constante e independencia de los residuos en un diseño factorial se verifican principalmente con los métodos gráficos presentados en el capítulo 3 para diseños con un solo factor. También se pueden aplicar los métodos analíticos descritos en ese capítulo.

Para el ejemplo 5.2, la independencia no la verificamos por no tener el orden en el cual se hicieron las corridas experimentales. En la figura 5.7 se grafican los residuos *vs.* predichos, y se observa que si se cumple el supuesto de varianza constante, al caer todos los puntos dentro de una banda horizontal. Asimismo, se cumple la normalidad al caer los residuos alineados en la gráfica de probabilidad normal. Otro

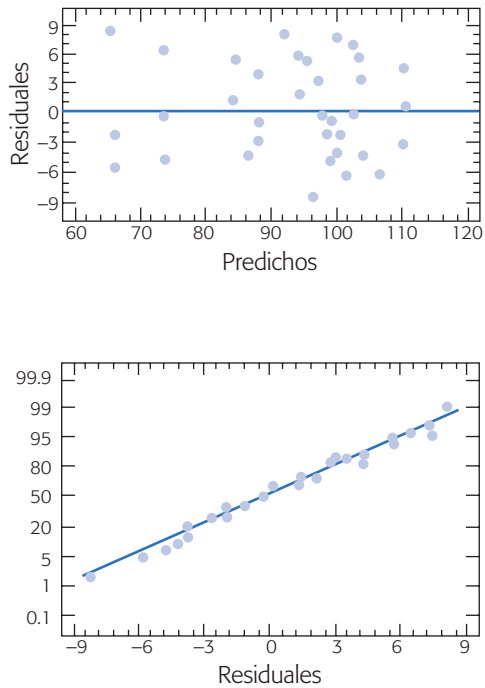


Figura 5.7 Predichos (Y_i) *vs*. residuos (e_i) y residuos en papel normal para el ejemplo 5.2.

aspecto importante a observar en estas gráficas es la ausencia de observaciones atípicas o aberrantes.

Diseños factoriales con tres factores

Cuando se quiere investigar la influencia de tres factores (A, B y C) sobre una o más variables de respuesta, y el número de niveles de prueba en cada uno de los factores es a, b y c, respectivamente, se puede construir el arreglo factorial $a \times b \times c$, que consiste de $a \times b \times c$ tratamientos o puntos experimentales. Entre los arreglos de este tipo que se utilizan con frecuencia en aplicaciones diversas se encuentran: el factorial 2^3, el factorial 3^3 y los factoriales mixtos con no más de cuatro niveles en dos de los factores, por ejemplo, el factorial $4 \times 3 \times 2$ y el factorial $4 \times 4 \times 2$, por mencionar dos de ellos.

Ejemplo 5.3

El experimento. Se desea investigar el efecto del tipo de suspensión (A), abertura de malla (B) y temperatura de ciclaje (C) en el volumen de sedimentación $Y(\%)$ de una suspensión. Para ello se decide correr un experimento factorial $3 \times 2 \times 2$ con seis réplicas, y las observaciones obtenidas en las 72 corridas experimentales se muestran en la siguiente tabla:

	A_1		A_2		A_3	
	B_1	B_2	B_1	B_2	B_1	B_2
C_1	60, 75, 75 86, 70, 70	67, 73, 73 67, 68, 68	62, 68, 65 76, 65, 65	71, 80, 80 72, 80, 80	76, 71, 75 70, 68, 73	75, 75, 75 75, 75, 77
C_2	55, 53, 53 55, 55, 55	52, 52, 57 52, 54, 54	44, 44, 45 48, 48, 45	60, 60, 60 67, 67, 65	52, 51, 50 52, 48, 54	56, 55, 57 59, 50, 55

Los niveles de prueba para cada factor, tanto en unidades originales como en unidades codificadas, se muestran en la siguiente tabla:

Factor	U. originales			U. codificadas		
	Bajo	Medio	Alto	Bajo	Medio	Alto
A: Tipo de suspensión	A_1	A_2	A_3	–1	0	1
B: Abertura de malla	40	–	60	–1	–	1
C: Temperatura	0	–	30	–1	–	1

La representación geométrica del experimento se muestra en la figura 5.8. Nótese que el factor *A* tiene tres niveles porque interesa evaluar precisamente tres suspensiones.

Modelo estadístico

En un diseño factorial $a \times b \times c$ como el del ejemplo, se supone que el comportamiento de la respuesta Y puede describirse mediante el modelo de efectos dado por:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl};$$

$$i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, b;\ k = 1, 2, \ldots, c;\ l = 1, 2, \ldots, n$$

donde μ es la media general, α_i es el efecto del nivel i-ésimo del factor A, β_j es el efecto del nivel j del factor B y γ_k es el efecto del nivel k en el factor C; $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$ y $(\beta\gamma)_{jk}$ representan efectos de interacción dobles (de dos factores) en los niveles ij, ik, jk, respectivamente, y $(\alpha\beta\gamma)_{ijk}$ es el efecto de interacción triple en la combinación o punto ijk; ε_{ijkl} representa el error aleatorio en la combinación $ijkl$ y l son las repeticiones o réplicas del experimento. Todos los efectos cumplen la restricción de sumar cero, es decir, son desviaciones relacionadas con la media general μ. De manera alternativa, se tiene el modelo de regresión dado por

$$Y_{ijkl} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2j} + \beta_3 X_{3k} + \beta_{12} X_{1i} X_{2j} + \beta_{13} X_{1i} X_{3k} + \beta_{23} X_{2j} X_{3k} + \beta_{123} X_{1i} X_{2j} X_{3k} + \varepsilon_{ijkl};$$
$$i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, b;\ k = 1, 2, \ldots, c;\ l = 1, 2, \ldots, n$$

Para mayores detalles con respecto a este modelo de regresión véase la sección "Regresión lineal múltiple" del capítulo 11.

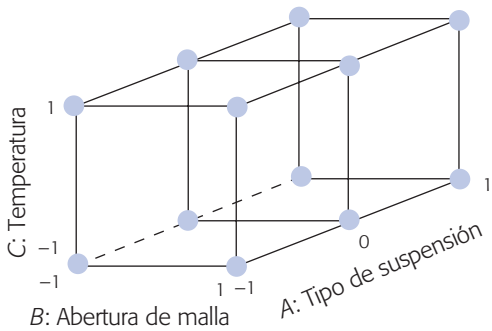


Figura 5.8 Representación geométrica del factorial 3 × 2 × 2 usado en el ejemplo 5.3.

Hipótesis de interés

El estudio factorial de tres factores (A, B y C) permite investigar los efectos: A, B, C, AB, AC, BC y ABC, donde el nivel de desglose o detalle con el que pueden estudiarse depende del número de niveles utilizado en cada factor. Por ejemplo, si un factor se prueba en dos niveles, todo su efecto marginal (individual) es lineal, o sea que su efecto individual no se puede descomponer; pero, si tuviera tres niveles su efecto marginal se puede descomponer en una parte lineal y otra cuadrática pura (véanse capítulos 7 y 12).

En resumen, se tienen siete efectos de interés sin considerar desglose, y con ellos se pueden plantear las siete hipótesis nulas H_0 : Efecto $A = 0$, H_0 : Efecto $B = 0$, …, H_0 : Efecto $ABC = 0$, cada una aparejada con su correspondiente hipótesis alternativa. El ANOVA para probar estas hipótesis se muestra en la tabla 5.6.

Al efecto cuyo valor-p sea menor al valor especificado para alfa, se declara estadísticamente significativo o se dice que está activo. Las sumas de cuadrados son muy similares a las obtenidas para dos factores (ecuaciones 5.2 y 5.3); habrá que considerar un subíndice adicional para el tercer factor, y comenzando otra vez por la suma total de cuadrados, éstas resultan ser:

$$SC_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} Y_{jkl}^2 - \frac{Y_{\cdots\cdot}^2}{N} \qquad \textbf{(5.5)}$$

donde $N = abcn$ es el total de observaciones en el experimento. Las sumas de cuadrados de efectos son

$$SC_A = \sum_{i=1}^{a} \frac{Y_{i\cdots}^2}{bcn} - \frac{Y_{\cdots\cdot}^2}{N};\ SC_B = \sum_{j=1}^{b} \frac{Y_{\cdot j\cdot\cdot}^2}{acn} - \frac{Y_{\cdots\cdot}^2}{N};\ SC_C = \sum_{k=1}^{c} \frac{Y_{\cdot\cdot k\cdot}^2}{abn} - \frac{Y_{\cdots\cdot}^2}{N};$$

$$SC_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{Y_{ij\cdot\cdot}^2}{cn} - \frac{Y_{\cdots\cdot}^2}{N} - SC_A - SC_B;$$

$$SC_{AC} = \sum_{i=1}^{a}\sum_{k=1}^{c} \frac{Y_{i\cdot k\cdot}^2}{bn} - \frac{Y_{\cdots\cdot}^2}{N} - SC_A - SC_C;$$

$$SC_{BC} = \sum_{j=1}^{b}\sum_{k=1}^{c} \frac{Y_{\cdot jk\cdot}^2}{an} - \frac{Y_{\cdots\cdot}^2}{N} - SC_B - SC_C;$$

Tabla 5.6 ANOVA para el diseño factorial $a \times b \times c$. En el ejemplo 5.3 es $3 \times 2 \times 2$.

FV	SC	GL	CM	F_0	Valor-p
Efecto A	SC_A	$a-1$	CM_A	CM_A/CM_E	$P(F > F_0^A)$
Efecto B	SC_B	$b-1$	CM_B	CM_B/CM_E	$P(F > F_0^B)$
Efecto C	SC_C	$c-1$	CM_C	CM_C/CM_E	$P(F > F_0^C)$
Efecto AB	SC_{AB}	$(a-1)(b-1)$	CM_{AB}	CM_{AB}/CM_E	$P(F > F_0^{AB})$
Efecto AC	SC_{AC}	$(a-1)(c-1)$	CM_{AC}	CM_{AC}/CM_E	$P(F > F_0^{AC})$
Efecto BC	SC_{BC}	$(b-1)(c-1)$	CM_{BC}	CM_{BC}/CM_E	$P(F > F_0^{BC})$
Efecto ABC	SC_{ABC}	$(a-1)(b-1)(c-1)$	CM_{ABC}	CM_{ABC}/CM_E	$P(F > F_0^{ABC})$
Error	SC_E	$abc(n-1)$	CM_E		
Total	SC_T	$abcn-1$			

$$SC_{ABC} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{Y_{ijk\bullet}^2}{n} - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{N} - SC_{AB} - SC_{AC} - SC_{BC}$$

Al restar éstas del total, la suma de cuadrados del error resulta ser

$$SC_E = SC_T - SC_A - SC_B - SC_C - SC_{AB} - SC_{AC} - SC_{BC} - SC_{ABC} \qquad \textbf{(5.6)}$$

cuyos respectivos grados de libertad se dan en la tabla 5.6. Una vez hecho el ANOVA, se procede a interpretar los efectos activos, y luego (aunque no necesariamente después) a diagnosticar la calidad del modelo.

El análisis de varianza para el caso del ejemplo 5.3 se muestra en la tabla 5.7. De aquí se concluye que no influyen los efectos *ABC*, *AC* ni *A*, dado que su valor-*p* es mayor que $\alpha = 0.05$. Por otra parte, se encuentran activos los efectos *B*, *C*, *AB* y en menor medida *BC*. Éstos son los cuatro efectos que se deben interpretar. Los efectos que no influyeron se pueden "eliminar" mandándolos al término error.[1] El ANOVA simplificado, pero con el efecto *A* se muestra en la tabla 5.8. Note que el CM_E en ambos ANOVA es prácticamente igual. En general se recomienda interpretar sólo los efectos significativos.

Interpretación de efectos activos. Del F_0 de la tabla 5.8 se aprecia que el efecto más importante es el de *C* seguido por *B* y la interacción de *AB*. En la figura 5.9 se muestran las gráficas de efectos de interacción *AB* y *BC*; se quiere minimizar el % de sedimentación. En el diagrama de la izquierda de la figura 5.9, el aspecto quebrado de las líneas indica que en el efecto de interacción *AB* está predominando la parte de curvatura sobre la parte lineal; la mejor combinación es la suspensión intermedia y abertura baja ($A = 0$, $B = -1$). En la parte derecha de la figura 5.9 las líneas se ven casi paralelas, lo cual es evidencia visual a favor de la poca importancia de la inter-

[1] Cuando se trabaja con el modelo de regresión debe cuidarse que el modelo final sea jerárquico, lo cual significa dejar en el mismo a los efectos simples que componen a interacciones de mayor orden. En este caso, por ejemplo, se dejaría en el modelo el efecto *A* aunque no influya (capítulo 12).

Tabla 5.7 ANOVA completo para el ejemplo 5.3.

FV	SC	GL	CM	F_0	Valor-*p*
A: Tipo	13.86	2	6.93	0.49	0.6126
B: Aber	480.5	1	480.5	34.25	0.0000
C: Temp	6 086.72	1	6 086.72	433.90	0.0000
AB	788.25	2	394.12	28.10	0.0000
AC	40.86	2	20.43	1.46	0.2412
BC	56.89	1	56.89	4.06	0.0485
ABC	31.03	2	15.51	1.11	0.3375
Error	841.66	60	14.03		
Total	8 339.78	71			

acción *BC*, y la fuerte pendiente de estas líneas tiene que ver con la influencia del factor *C* : *temp*. En otras palabras, en la gráfica del efecto *BC* se observa prácticamente sólo el efecto de *C*, por ello, no es necesario representar a este último. Para minimizar la mejor combinación para el factor *C* es su nivel alto. En resumen, el mejor tratamiento es ($A = 0$, $B = -1$, $C = 1$).

Diagnóstico. Es importante evaluar, mediante el análisis de residuos, la calidad del modelo de efectos antes de considerar la combinación (0, –1, 1) como la mejor. De antemano podemos afirmar que la contundencia del valor-*p* en la tabla de ANOVA (tabla 5.8) es tan fuerte que es muy difícil que una violación moderada de los supuestos del modelo cambie la conclusión obtenida. Sin embargo, siempre que sea posible debe comprobarse el cumplimiento de los supuestos de normalidad, varianza constante, independencia y la ausencia de observaciones atípicas. Las gráficas de residuos contra predichos y en papel normal se muestran en las figuras 5.10*ab*.

Se observa en ellas una aparente violación al supuesto de varianza constante y de normalidad. Pero otra posibilidad es la presencia de tres observaciones alejadas, las cuales se asocian con dos residuos grandes y uno pequeño que hacen ver el no cumplimiento de los supuestos. Trate de visualizar las gráficas ignorando esos puntos y verá que los supuestos se cumplirían. Aquí tiene sentido pensar en tres observaciones alejadas porque el experimento consiste en 72 observaciones. Valdría la pena investigar por lo menos si dichas observaciones tienen algo en común.

Tabla 5.8 ANOVA simplificado para el ejemplo 5.3.

FV	SC	GL	CM	F_0	Valor-*p*
A: Tipo	13.86	2	6.93	0.49	0.6126
B: Aber	480.5	1	480.5	33.66	0.0000
C: Temp	6 086.72	1	6 086.72	426.41	0.0000
AB	788.25	2	394.12	27.61	0.0000
BC	56.89	1	56.89	3.99	0.0502
Error	913.55	64	14.27		
Total	8 339.78	71			

Si se admite que tales observaciones son parte del proceso de sedimentación que se estudia, entonces se tiene el no cumplimiento de los supuestos de varianza constante y normalidad. De hecho, la prueba Shapiro Wilks rechaza que los residuos sean normales (valor-p = 0.002). El no cumplimiento de la igualdad de varianzas es

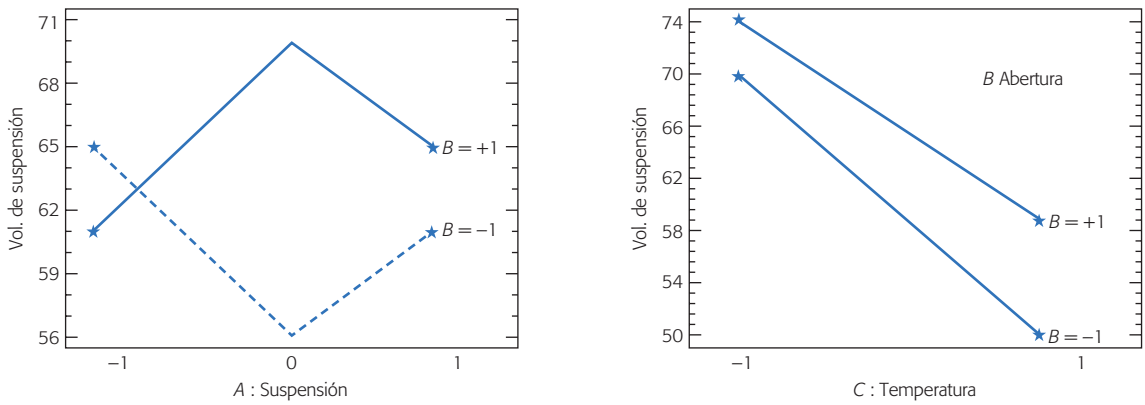


Figura 5.9 Efectos de interacción *AB* y *BC* en el ejemplo 5.3.

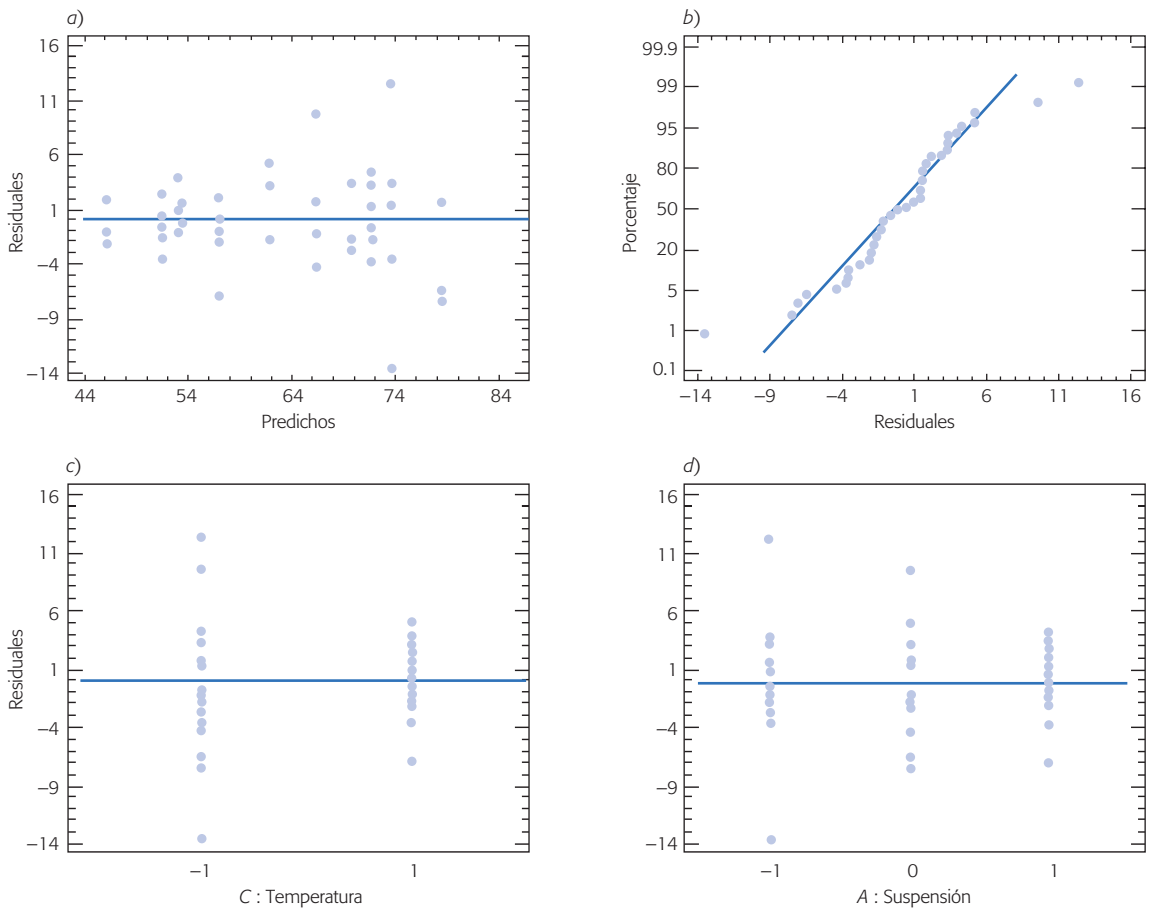


Figura 5.10 Gráficas de residuos: *a*) residuos *vs.* predichos, *b*) residuos en papel normal, *e*) y *d*) residuos *vs.* factores *A* y *C* para el ejemplo 5.3.

posible que se deba a que los factores A : *Suspen* y C : *Temp* tienen un efecto real sobre la dispersión del porcentaje de sedimentación, de acuerdo a como se observa en las figuras 5.11*cd*. En caso de existir tal efecto, debe tomarse en cuenta para también buscar cómo reducir la variabilidad de la respuesta.

A pesar de lo antes dicho, y dado el valor-p contundente en el ANOVA (tabla 5.8), se prevé la no afectación en la conclusión sobre el mejor tratamiento. Sin embargo, para estar seguros se intenta corregir el problema de violación de supuestos analizando una transformación apropiada de la respuesta Y, como se hace a continuación.

Transformaciones para estabilizar varianza

En la práctica, algunas variables de respuesta no siguen una distribución normal sino que se distribuyen, por ejemplo *Poisson*, *binomial* o *Gamma*, por mencionar tres casos. Resulta que en estas distribuciones la media está relacionada con la desviación estándar (variabilidad) y, naturalmente, al cambiar la media de un tratamiento a otro, con ella cambia la variabilidad de la respuesta. También es cierto que al suponer normalidad y varianza constante, éstas no se tienen que cumplir de manera estricta, dado que el procedimiento de ANOVA es robusto o admite desviaciones moderadas de dichos supuestos.

Existen al menos tres maneras de solucionar o minimizar el problema por falta de normalidad y de varianza heterogénea en los residuos: 1) utilizar métodos de análisis no paramétricos, que no requieren las suposiciones de normalidad y varianza constante (véase Conover, 1980); 2) hacer el análisis mediante modelos lineales generalizados (GLM), en los que se ajusta un modelo lineal usando otras distribuciones diferentes a la normal, donde la varianza no tiene por qué ser constante (Myers, *et al*., 2002), y 3) hacer el análisis sobre la respuesta transformada a una escala en la que los supuestos se cumplan. En este libro sólo se presenta el tercer enfoque.

La transformación más apropiada de la respuesta para corregir o minimizar los problemas de falta de normalidad y de varianza constante, depende del tipo de relación que existe entre la media y la varianza de Y. Esta relación se puede visualizar en la gráfica de residuos *vs*. predichos (véase figura 5.10*a*). Según lo pronunciado que sea la "forma de corneta" de los puntos en dicha gráfica, se determina la transformación más apropiada. Con un paquete estadístico se pueden probar varias transformaciones para elegir aquella en la cual los supuestos se cumplan de mejor manera. Las transformaciones más usuales se muestran en la tabla de la página siguiente.

En la tabla el símbolo $\propto$ significa "es proporcional a". A medida que se da la relación de proporcionalidad con respecto a mayor potencia de la media o valor esperado, se requiere una transformación más fuerte para lograr igualdad de varianzas en el análisis de la respuesta transformada. El grado de proporcionalidad se puede ver en la gráfica de *residuos vs. predichos*.

Ejemplo 5.4

En la figura 5.10, correspondiente al ejemplo 5.3, vimos ligeras violaciones de los supuestos de normalidad y varianza constante de los residuos del modelo ajustado en el ANOVA. Para corregir este problema se aplicó la transformación $Y' = 1/Y$

Relación entre la varianza σ^2 y la media ($E(Y)$)	Transformación apropiada	Tipo de transformación
$\sigma^2 \propto E(Y)(1-E(Y))$	$Y' = \text{sen}^{-1}(\sqrt{Y})$	Arcoseno, útil cuando la respuesta Y son proporciones (se distribuye binomial)
$\sigma^2 \propto E(Y)$	$Y' = \sqrt{Y}$	Raíz cuadrada, para los datos tipo Poisson
$\sigma^2 \propto [E(Y)])^2$	$Y' = \ln(Y)$ o $Y' = \log_{10}(Y)$	Transformación logaritmo
$\sigma^2 \propto [E(Y)]^3$	$Y' = Y^{-1/2}$	Recíproco de la raíz cuadrada
$\sigma^2 \propto [E(Y)]^4$	$Y' = Y^{-1}$	Recíproco

(recíproca), con los datos transformados se obtuvo el ANOVA de la tabla 5.9. Observe que el valor-p del efecto BC indica claramente que éste es estadísticamente significativo, pero se puede mostrar a través de F_0 que al igual que antes, su contribución en la explicación de la respuesta es escasa.

Debe tenerse cuidado al interpretar las gráficas de efectos en la variable transformada $Y' = 1/Y$ puesto que los valores grandes de Y corresponden a los más pequeños de Y' y los valores originalmente pequeños ahora son los grandes.

Las gráficas de residuos obtenidos del modelo con transformación se muestran en la figura 5.11. Si se compara esta figura con la que se tenía para el modelo sin usar transformación (figura 5.10), se nota cómo la validez de los supuestos se cumple mejor.

Cabe aclarar que las transformaciones para estabilizar la varianza no eliminan el efecto de dispersión que de por sí existe. Sólo permiten analizar mejor el efecto sobre la media.

Diseño factorial general

Lo que se ha dicho para los dos diseños factoriales con 2 y 3 factores puede extenderse fácilmente para cuando se tienen más factores. Considere f factores A, B, C, …, K con niveles a, b, c, …, k, respectivamente, donde la letra K denota al f-ésimo o último factor del conjunto a estudiar, no necesariamente el undécimo, que es el lugar de esta letra en el alfabeto. Con estos niveles y factores se puede construir el diseño factorial general $a \times b \times \ldots \times k$, que consiste de $a \times b \times \ldots \times k$ tratamientos o puntos de prueba. Con este diseño se pueden estudiar f efectos principales, $f(f-1)/2$ interacciones dobles, $f(f-1)(f-2)/(3 \times 2)$ interacciones triples, y así sucesivamente hasta la única interacción de los f factores ($ABC \ldots K$). El cálculo del número de interacciones de cierta cantidad m de factores se hace mediante la operación "combinaciones de f en m" $\left(\binom{f}{m} = \dfrac{f!}{f!(m!(f-m)!}\right)$ que cuenta el número de diferentes maneras de seleccionar m factores de los f, donde $f! = f \times (f-1) \times \ldots \times 2 \times 1$.

Por ejemplo, el diseño factorial 2^5 tiene cinco efectos principales, 10 interacciones dobles, 10 interacciones triples, cinco interacciones cuádruples y una interacción quíntuple, lo cual da un total de 31 efectos. Por su parte, el factorial 3^5 también tiene este mismo número de efectos, pero al contar con tres niveles en cada factor,

Tabla 5.9 ANOVA para el ejemplo 5.4, respuesta $Y^{-1/2}$.

FV	SC	GL	CM	F_0	Valor-*p*
A: Tipo	7.41E-7	2	3.71E-7	0.40	0.6724
B: Aber	0.000043	1	0.000043	46.72	0.0000
C: Temp	0.00042	1	0.00042	455.15	0.0000
AB	0.000062	2	0.000031	33.38	0.0000
BC	0.000013	1	0.000013	13.98	0.0004
Error	0.000059	64	9.28E-7		
Total	0.00060	71			

cada efecto principal se puede descomponer en su parte lineal y cuadrática (véase capítulo 7). Cabe destacar que mientras el diseño factorial 2^5 tiene 32 tratamientos, el factorial 3^5 tiene 243, una cantidad de tratamientos difícil de manejar. Aun si pudiera correrse, representa una opción muy ineficaz; además, existen arreglos experimentales más pequeños y eficientes (capítulo 8).

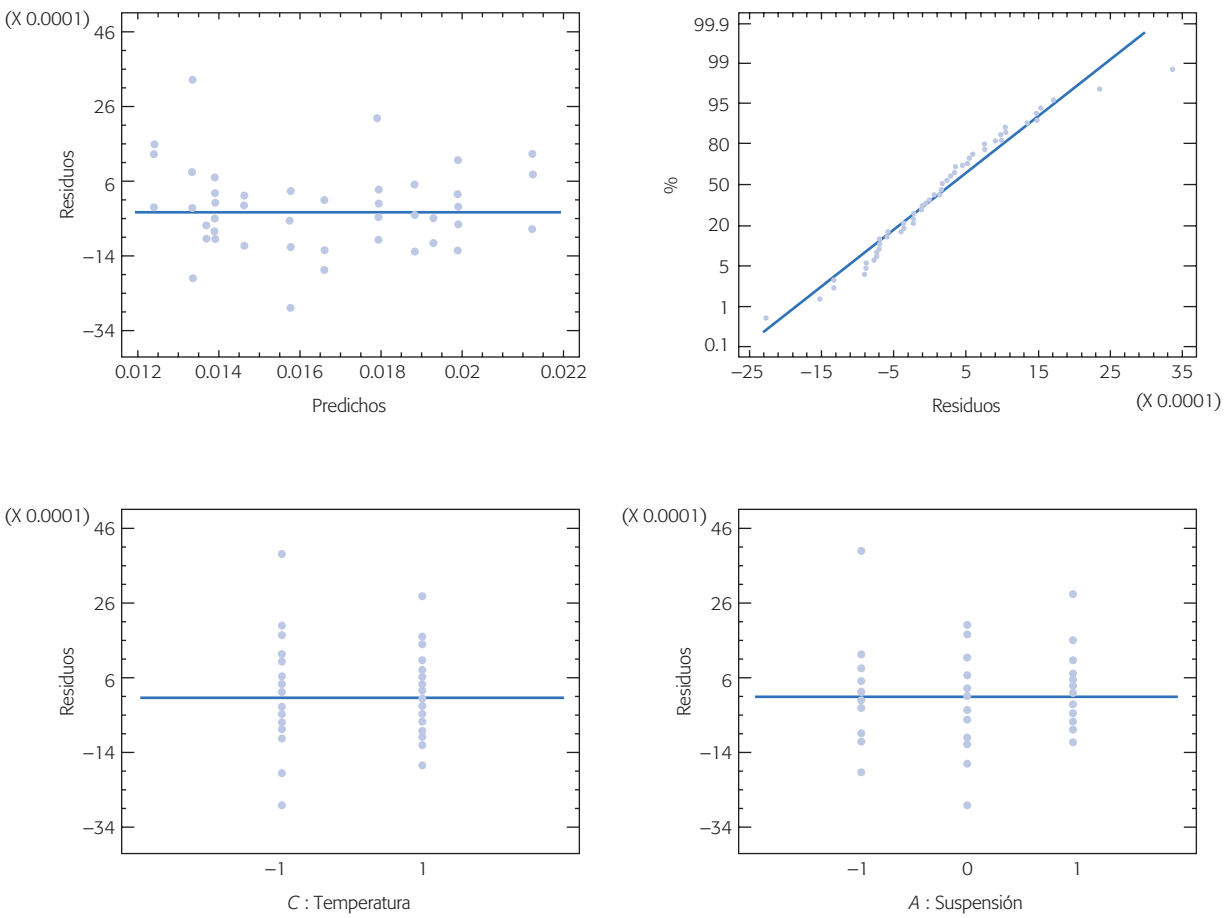


Figura 5.11 Gráficas de residuales obtenidos del análisis de la respuesta l/*Y*, ejemplo 5.4.

De acuerdo con lo antes dicho, en el factorial general $a \times b \times \ldots \times k$, se pueden plantear $2^f - 1$ hipótesis que se prueban mediante el análisis de varianza. Si se tienen n réplicas. Las primeras tres columnas de este ANOVA se muestran en la tabla 5.10.

La suma de cuadrados totales está dada por:

$$SC_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\cdots\sum_{m=1}^{k}\sum_{r=1}^{n} Y_{ij\cdots mr}^2 - \frac{Y_{\bullet\bullet\cdots\bullet\bullet}^2}{N}$$

donde $N = abc \ldots kn$ es el total de observaciones en el experimento. Las sumas de cuadrados de efectos son:

$$SC_A = \sum_{i=1}^{a} \frac{Y_{i\bullet\cdots\bullet}^2}{bc\cdots lkn} - \frac{Y_{\bullet\bullet\cdots\bullet}^2}{N};\ SC_B = \sum_{j=1}^{b} \frac{Y_{\bullet j\cdots\bullet}^2}{ac\cdots lkn} - \frac{Y_{\bullet\bullet\cdots\bullet}^2}{N};$$

$$SC_C = \sum_{k=1}^{c} \frac{Y_{\bullet\bullet k\cdots\bullet}^2}{ab\cdots lkn} - \frac{Y_{\bullet\bullet\cdots\bullet}^2}{N};\cdots;\ SC_K = \sum_{m=1}^{k} \frac{Y_{\bullet\bullet\cdots\bullet m\bullet}^2}{abc\cdots ln} - \frac{Y_{\bullet\bullet\cdots\bullet}^2}{N};$$

$$SC_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{Y_{ij\bullet\cdots\bullet}^2}{c\cdots lkn} - \frac{Y_{\bullet\bullet\cdots\bullet}^2}{N} - SC_A - SC_B$$

$$SC_{(K-1)K} = \sum_{p=1}^{l}\sum_{m=1}^{k} \frac{Y_{\bullet\bullet\cdots pm\bullet}^2}{abc\cdots n} - \frac{Y_{\bullet\bullet\cdots\bullet}^2}{N} - SC_{(K-1)} - SC_K;$$

$$SC_{ABC} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} \frac{Y_{ijk\cdots\bullet}^2}{\ldots lkn} - \frac{Y_{\bullet\bullet\cdots\bullet}^2}{N} - SC_{AB} - SC_{AC} - SC_{BC}$$

$$SC_{AB\cdots K} = \sum_{i=1}^{a}\sum_{j=1}^{b}\cdots\sum_{p=1}^{l}\sum_{m=1}^{k} \frac{Y_{ij\cdots pm\bullet}^2}{n} - \frac{Y_{\bullet\bullet\cdots\bullet}^2}{N} - SC_A - \cdots - SC_{AB} - \cdots - SC_{AB\cdots K-1}$$

Tabla 5.10 ANOVA para el diseño factorial general $a \times b \times \ldots \times k$.

FV	SC	GL
$Ef.\ A$	SC_A	$a-1$
⋮	⋮	⋮
$Ef.\ K$	SC_K	$k-1$
$Ef.\ AB$	SC_{AB}	$(a-1)(b-1)$
⋮	⋮	⋮
$Ef.\ K(K-1)$	$SC_{(K-1)K}$	$(l-1)(k-1)$
$Ef.\ ABC$	SC_{ABC}	$(a-1)(b-1)(c-1)$
⋮	⋮	⋮
$Ef.\ (K-2)(K-1)K$	$SC_{(K-2)(K-1)K}$	$(m-1)(l-1)(k-1)$
⋮	⋮	⋮
$Ef.\ AB\cdots K$	$SC_{AB\cdots K}$	$(a-1)(b-1)\cdots(k-1)$
Error	SC_E	$abc\cdots k(n-1)$
Total	SC_T	$(abc\cdots kn)-1$

Al final, la suma de cuadrado del error se calcula por sustracción,

$$SC_E = SC_T - SC_A - \cdots - SC_K - SC_{AB} - \cdots - SC_{(K-1)K} - SC_{ABC} - \cdots - SC_{AB\cdots K} \tag{5.7}$$

En el ANOVA de la tabla 5.10 para el factorial general $a \times b \times \ldots \times k$ se observa la necesidad de contar con al menos dos réplicas del experimento para calcular la suma de cuadrados del error (SC_E), y completar toda la tabla de ANOVA. Sin embargo, esta necesidad de réplicas ($n = 2$), que se ha mencionado a lo largo del capítulo, es para el caso irreal de que interesaran los $2^f - 1$ efectos. Pero resulta que, con excepción del factorial 2^2, en un factorial completo prácticamente nunca interesan todos sus posibles efectos, puesto que en términos generales sólo algunos de ellos están activos. El *principio de Pareto*, que en este contexto también se llama principio de *esparcidad de efectos*, dice que la mayoría de la variabilidad observada se debe a unos pocos de los efectos posibles; por lo común se debe a algunos efectos principales e interacciones dobles.

Principio de Pareto
Se refiere a que la mayoría de la variabilidad observada se debe a unos pocos de los efectos posibles.

Modelos de efectos aleatorios

Hasta aquí los modelos de efectos que se han utilizado son *modelos de efectos* o *factores fijos*, lo cual significa que todos los niveles de prueba en cada factor son todos los disponibles para ese factor, o bien, se estudian todos los niveles de interés en ese factor; es en este sentido que los niveles están fijos. Éste es el caso, por ejemplo, cuando en el factor operador se toman los tres únicos operadores como los niveles de prueba, o cuando los niveles del factor máquinas son las cuatro máquinas existentes. O bien, cuando se comparan tres tipos de material porque son los que interesa comparar aunque existan otros materiales de ese tipo. Con factores fijos, las conclusiones obtenidas sólo son válidas para los niveles de prueba que se estudian en el experimento.

Factor fijo
Se refiere a que los niveles de prueba en un factor son todos los niveles disponibles para éste.

En ocasiones, los niveles de prueba son una muestra aleatoria de la población de niveles posibles. En este caso es más apropiado utilizar un modelo de *efectos* o *factores aleatorios*. Un ejemplo de esta situación es cuando se prueban cinco instrumentos de medición, pero la población de los mismos es de 100 instrumentos; obviamente, no es posible experimentar con todos los equipos. Entonces se experimenta sólo con cinco de ellos elegidos al azar, y las conclusiones obtenidas se infieren como válidas para la población entera de instrumentos.

Factor aleatorio
Cuando los niveles de prueba utilizados en un factor son una muestra aleatoria de la población de niveles para ese factor.

La aplicación de un modelo de efectos aleatorios conlleva la necesidad de considerar la incertidumbre asociada con la elección aleatoria de los niveles de prueba. Es decir, ya no tiene sentido, para un factor A, preocuparse por el efecto α_i del nivel i como con efectos fijos. Lo que ahora (con efectos aleatorios) tiene sentido es hablar de la varianza con la que el factor aleatorio contribuye a la variación total; es decir, es preciso estimar dicha varianza y probar si su contribución a la variabilidad total es significativa.

El caso de dos factores aleatorios

Si se consideran dos factores aleatorios A y B, de los cuales se prueban a y b niveles elegidos de una población grande de niveles, entonces si los $a \times b$ tratamientos se replican n veces, el modelo de efectos aleatorios es

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}; \quad i = 1, 2, \ldots, a; \quad j = 1, 2, \ldots, b; \quad k = 1, 2, \ldots, n$$

donde μ es la media general, α_i es el efecto debido al i-ésimo nivel del factor A, β_j es el efecto del j-ésimo nivel del factor B, $(\alpha\beta)_{ij}$ representa al efecto de interacción en la combinación ij y ε_{ijk} es el error aleatorio que se supone sigue una distribución normal con media cero y varianza constante, $N(0, \sigma^2)$ y son independientes entre sí. El aspecto de este modelo es igual al de efectos fijos (ecuación 5.1), pero el hecho de que los efectos sean aleatorios implica que no tiene sentido probar hipótesis directamente sobre tales efectos (medias), sino que ahora el interés se enfoca en estudiar la varianza de dichos efectos. Para ello, se supone que los términos α_i, β_j, $(\alpha\beta)_{ij}$ y ε_{ijk} son variables aleatorias independientes normales, con media cero y varianzas σ^2_α, σ^2_β, $\sigma^2_{\alpha\beta}$ y σ^2, respectivamente.

De esta manera, si se calcula la varianza en ambos lados del modelo anterior, se obtiene el modelo de componentes de varianza dado por:

$$\text{Var}(Y_{ijk}) = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2$$

Componentes de varianza
Son las contribuciones de cada efecto a la variación total en el modelo de efectos aleatorios.

donde σ^2_α, σ^2_β, $\sigma^2_{\alpha\beta}$ son las contribuciones de cada efecto a la variación total y se llaman *componentes de varianza*; σ^2 es el componente de varianza debido al error aleatorio. Las hipótesis de interés son $H_0 : \sigma^2_\alpha = 0$, $H_0 : \sigma^2_\beta = 0$ y $H_0 : \sigma^2_{\alpha\beta} = 0$. Los cálculos necesarios para probar estas hipótesis involucran las mismas sumas de cuadrados del modelo de efectos fijos (véase la sección "Diseños factoriales con dos factores" de este capítulo), de las cuales se obtienen los correspondientes cuadrados medios. Para obtener los estadísticos de prueba F_0 apropiados debe tomarse en cuenta que los valores esperados de los cuadrados medios son

$$\begin{aligned} E(CM_A) &= \sigma^2 + n\sigma^2_{\alpha\beta} + bn\sigma^2_\alpha \\ E(CM_B) &= \sigma^2 + n\sigma^2_{\alpha\beta} + an\sigma^2_\beta \\ E(CM_{AB}) &= \sigma^2 + n\sigma^2_{\alpha\beta} \\ E(CM_E) &= \sigma^2 \end{aligned}$$

de tal forma que para probar las hipótesis $H_0 : \sigma^2_\alpha = 0$, $H_0 : \sigma^2_\beta = 0$ y $H_0 : \sigma^2_{\alpha\beta} = 0$, *los estadísticos de prueba apropiados*[2] en el ANOVA son $F_0^A = CM_A/CM_{AB}$, $F_0^B = CM_B/CM_{AB}$ y $F_0^{AB} = CM_{AB}/CM_E$, respectivamente. Observe que en el modelo de

[2] El estadístico apropiado para una hipótesis específica se determina encontrando dos cuadrados medios que al dividir sus valores esperados, y suponiendo H_0 verdadera, el cociente es uno. Por ejemplo, CM_A/CM_{AB} es el estadístico correcto para la hipótesis $H_0 : \sigma^2_\alpha = 0$ porque $E(CM_A)/E(CM_{AB}) = 1$ cuando $\sigma^2_\alpha = 0$.

efectos aleatorios los cuadrados medios de los efectos principales se comparan con el cuadrado medio de la interacción, y no con el cuadrado medio del error, como se hace en el modelo de efectos fijos. En caso de rechazar alguna de las hipótesis sobre las varianzas, se concluye que el efecto correspondiente contribuye de manera significativa a la variación de la respuesta. La conclusión práctica no consiste en determinar el mejor tratamiento, sino que generalmente se traduce en tomar medidas para que la contribución del componente de varianza se reduzca.

Al resolver las ecuaciones dadas por los valores esperados de cuadrados medios para los componentes de varianza, se obtienen estimadores de éstos en función de los cuadrados medios del error, esto es,

$$\begin{aligned} \hat{\sigma}^2 &= CM_E \\ \hat{\sigma}^2_{\alpha\beta} &= \frac{CM_{AB} - CM_E}{n} \\ \hat{\sigma}^2_{\alpha\beta} &= \frac{CM_B - CM_{AB}}{an} \\ \hat{\sigma}^2_{\alpha} &= \frac{CM_A - CM_{AB}}{bn} \end{aligned} \tag{5.8}$$

Ejemplo 5.5

Estudio R&R. En una compañía dedicada a la fabricación de bombas y válvulas, algunos componentes críticos tienen tolerancias muy estrechas que son difíciles de cumplir. De aquí que sea necesario estimar el error de medición con el fin de ver la posibilidad de reducirlo para cumplir con las especificaciones. El ancho de una pieza particular es una característica de calidad crítica, cuyas especificaciones son 69 ± 0.4 mm. Se eligen dos inspectores al azar y siete piezas para correr un experimento (estudio R&R), a fin de estimar la contribución de los inspectores, de las piezas y del error aleatorio (repetibilidad) en la variabilidad total observada. El experimento utilizado se muestra en la siguiente tabla:

Número de piezas	Inspector *Z*		Inspector *W*	
	1	2	1	2
1	69.38	69.60	69.62	69.52
2	39.72	69.80	69.78	69.90
3	69.58	69.70	69.70	69.92
4	69.50	69.50	69.46	69.50
5	69.48	69.40	69.50	69.42
6	69.56	69.40	69.68	69.64
7	69.90	70.02	69.94	69.88

Nótese que cada inspector mide dos veces cada pieza. Sean los inspectores el factor *A* y las piezas el factor *B*, el primero con dos niveles y el segundo con siete niveles, en ambos casos seleccionados al azar. El modelo de componentes de varianza

propuesto para describir estos datos es donde σ^2_α es el componente de varianza de los inspectores, σ^2_β es el componente debido a las piezas, $\sigma^2_{\alpha\beta}$ es el componente de interacción de ambos factores y σ^2 es el componente aleatorio. Interesa probar las hipótesis $H_0 : \sigma^2_\alpha = 0$, $H_0 : \sigma^2_\beta = 0$ y $H_0 : \sigma^2_{\alpha\beta} = 0$, y estimar los componentes de varianza. El ANOVA para probar estas hipótesis se muestra en la siguiente tabla:

FV	SC	GL	CM	F_0	Valor-*p*
A: Insp	0.00036	1	0.00036	0.069	0.8043
B: Pieza	0.7516	6	0.1252	24.07	0.0000
AB	0.0313	6	0.0052	0.75	0.6169
Error	0.097	14	0.0069		
Total	0.8803	27			

Las tres primeras columnas se obtienen igual que el modelo de efectos fijos, pero las dos últimas deben corregirse de acuerdo con el estadístico de prueba apropiado para un modelo de efectos aleatorios ($F_0^A = CM_A/CM_{AB} = 0.069$ y $F_0^B = CM_B/CM_{AB} = 24.07$). Los valor-*p* indican que la variabilidad de las piezas es estadísticamente diferente a cero, mientras que la variabilidad de los inspectores y de la interacción *inspector* × *pieza* no es significativa (es igual a cero). Desde el punto de vista del objetivo del experimento, los resultados del ANOVA son los deseados: la reproducibilidad ($\sigma^2_\alpha + \sigma^2_{\alpha\beta}$) es estadísticamente igual a cero, es decir, los inspectores no afectan el proceso de medición. La estimación de los componentes de varianza, a partir de los cuadrados medios, queda como:

$$\hat{\sigma}^2 = CM_E = 0.0069$$

$$\hat{\sigma}^2_{\alpha\beta} = \frac{CM_{AB} - CM_E}{n} = \frac{0.0052 - 0.0069}{2} = 0.000$$

$$\hat{\sigma}^2_\beta = \frac{CM_B - CM_{AB}}{an} = \frac{0.1252 - 0.0052}{2 \times 2} = 0.030$$

$$\hat{\sigma}^2_\alpha = \frac{CM_A - CM_{AB}}{bn} = \frac{0.0036 - 0.0052}{6 \times 2} = 0.000$$

De aquí se concluye que la reproducibilidad ($\sigma^2_\alpha + \sigma^2_{\alpha\beta} = 0$) no tiene contribución y la repetibilidad expresada como $5.15\hat{\sigma}$ es igual a 0.428. Si este valor se compara con la tolerancia de 0.8, se encuentra que ocupa 53% de ésta, cuando lo deseable es que este porcentaje sea menor al 10%, por lo que el instrumento es inadecuado para discriminar entre piezas buenas y malas. (Para mayores detalles sobre estudios R&R, consulte el capítulo 11 de Gutiérrez Pulido y De la Vara, 2004.)

Modelo mixto: factores aleatorios y fijos

En estos experimentos se tienen factores aleatorios y factores fijos. Por ejemplo, si el factor *A* es aleatorio y *B* es fijo, el modelo de componentes de varianza es:

$$\text{Var}(Y_{ijk}) = \sigma^2_\alpha + \sigma^2_{\alpha\beta} + \sigma^2$$

donde el componente individual del factor B desaparece, dado que es fijo, pero se mantiene el componente de interacción por ser A aleatorio. De manera que las hipótesis de interés que involucran al factor A son $H_0 : \sigma_\alpha^2 = 0$ y $H_0 : \sigma_{\alpha\beta}^2 = 0$, y para el factor B se prueba la misma hipótesis sobre medias de niveles del modelo de efectos fijos. En caso de rechazarla se determina cuál tratamiento o nivel del factor B es mejor. En el *modelo de efectos mixtos* la esperanza de los cuadrados medios es

Modelo de efectos mixtos
Son experimentos en los que se consideran factores aleatorios y factores fijos.

$$
\begin{aligned}
E(CM_A) &= \sigma^2 + bn\sigma_\alpha^2 \\
E(CM_B) &= \sigma^2 + n\sigma_{\alpha\beta}^2 + \frac{an\sum_{i=1}^{a}\beta_i^2}{b-1} \\
E(CM_{AB}) &= \sigma^2 + n\sigma_{\alpha\beta}^2 \\
E(CM_E) &= \sigma^2
\end{aligned}
$$

de aquí, los estadísticos apropiados para los efectos aleatorios son $F_0^A = CM_A/CM_E$ y $F_0^{AB} = CM_{AB}/CM_E$ y para los efectos fijos de B el estadístico adecuado es $F_0^B = CM_B/CM_{AB}$. Los componentes de varianza estimados son:

$$
\begin{aligned}
\hat{\sigma}^2 &= CM_E \\
\hat{\sigma}_{\alpha\beta}^2 &= \frac{CM_{AB} - CM_E}{n} \\
\hat{\sigma}_\alpha^2 &= \frac{CM_A - CM_E}{bn}
\end{aligned}
$$

Cómo hacerlo con software

Casi cualquier software estadístico incluye procedimientos para realizar el análisis de un experimento factorial. En particular, en *Statgraphics* el análisis de los diseños factoriales donde hay al menos uno que tiene más de dos niveles, se obtiene con la siguiente secuencia de opciones: *Special* → *Experimental Design* → *Create Design*. De la versión 15 en adelante de este software se usa la secuencia: *Doe* → *Design Creation* → *Create New Design*, después de esto se debe elegir el tipo de diseño, entre los cuales están *Multi-Factor Categorical* o *Multilevel-Factor*. El primero se usa si al menos un factor es categórico, o si aunque todos sean factores numéricos se desea un análisis de varianza no desglosado, como los que hemos visto en este capítulo. Por el contrario, si todos los factores son numéricos y se desea un análisis desglosado, tipo superficie de respuesta (véase capítulo 12), entonces se debe elegir la opción *Multilevel-Factor*.

En el caso del software Minitab, la secuencia es *Stat* → *DOE* → *Factorial* → *Create Factorial Design* y luego se elige el *General full Factorial Design*.

En cualquier caso, a continuación se eligen el número de factores y las variables de respuesta, se dan sus nombres y su número de niveles. Se define la variable de respuesta y después se da el número de réplicas, y si está haciendo una aplicación es

preciso asegurarse de activar la aleatorización (*randomize*). Lo anterior generará columnas para cada factor y para cada variable de respuesta.

Para hacer el análisis una vez que se haya generado el archivo de datos con los tratamientos y las respuestas, se sigue una secuencia similar a las indicadas en cada paso, pero ahora eligiendo las opciones de *Analyze*, donde se tendrá acceso a un conjunto de opciones de análisis tanto gráficas como analíticas, entre ellas las que se han expuesto en este capítulo.

Uso de Excel

El ANOVA de un diseño factorial con dos factores se realiza con la secuencia: *Herramientas* → *Análisis de datos* → *Análisis de dos factores con varias muestras por grupo*. Después se declara el rango de los datos, que pueden estar acomodados por columnas o por renglones. La salida contiene las estadísticas básicas de cada una de las muestras y el ANOVA correspondiente. Sin embargo, para más de dos factores, Excel no incluye un procedimiento predefinido.

Preguntas y ejercicios

1. ¿Qué es un experimento factorial completo?
2. ¿Cuántos efectos se pueden estudiar con un factorial 4 × 3 × 2? Bosqueje su tabla de análisis de varianza.
3. Mencione al menos tres ventajas de la experimentación factorial sobre la estrategia de mover un factor a la vez.
4. ¿Cuál es la implicación práctica de utilizar tres niveles de prueba en lugar de dos en un factor dado?
5. ¿Por qué no tiene sentido utilizar el modelo de regresión cuando los factores son cualitativos? Si fueran cuantitativos, ¿qué se gana con el modelo de regresión en relación al modelo de efectos?
6. ¿Cómo se construye la gráfica de un efecto de interacción doble? ¿Cómo se interpreta?
7. ¿Cuáles son los supuestos del modelo en un diseño factorial y con cuáles gráficas de residuos se puede verificar cada uno de estos supuestos?
8. En la pregunta anterior, ¿cómo se vería en las gráficas un punto muy alejado o aberrante?
9. De los tres supuestos del modelo, ¿cuál puede afectar más el análisis en caso de no cumplirse?
10. En caso de no cumplirse los supuestos de normalidad y varianza constante, ¿qué se puede hacer para evitar problemas con el análisis y los resultados obtenidos?
11. ¿Con base en qué se puede encontrar una transformación adecuada de la respuesta cuando no se cumplen los supuestos?
12. ¿Qué significa que el modelo estadístico sea de efectos aleatorios? ¿En qué cambian las hipótesis de interés en factor aleatorio con respecto de uno fijo?
13. Represente en el plano cartesiano un diseño factorial 4 × 4.

14. A continuación se muestra parte del ANOVA para un diseño factorial 3 × 5 con dos réplicas, el factor *A* con tres niveles y el *B* con cinco.

Factores de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Razón *F*
A	800			
B	900			
AB	300			
Error	400			
Total				

a) Suponga efectos fijos, anote el modelo estadístico apropiado y formule las hipótesis a probar para este experimento.
b) Agregue en esta tabla los grados de libertad, el cuadrado medio y la razón *F* para cada una de las fuentes de variación.
c) Explique de manera esquemática cómo calcularía el valor-*p* para *A*, por ejemplo.
d) Con la información disponible ¿se pueden hacer conjeturas sobre cuáles de las fuentes de variación son significativas estadísticamente? Argumente su respuesta.

15. Conteste todo el ejercicio anterior, pero ahora suponiendo que ambos factores son aleatorios.

16. Conteste todos los incisos del ejercicio 14, pero ahora suponga que el factor *A* es fijo y el factor *B* es aleatorio.

17. Cuando se hace un ANOVA de un diseño factorial del tipo $a \times b$ con n réplicas, se detecta que ambos factores influyen sobre la variable de respuesta, entonces surge la necesidad de comparar las medias de los niveles de cada factor para ver cuáles de ellas son diferentes entre sí. Conteste:
a) ¿Cuál es la expresión para la diferencia mínima significativa (LSD) para comparar los niveles del factor *B*, sin tomar en cuenta el posible efecto de interacción?
b) Conteste el inciso anterior pero ahora para el factor *A*.
c) Conteste los dos incisos anteriores, pero ahora tome en cuenta el efecto de interacción.
d) ¿Cuándo se espera que ambas formas de proceder arrojen conclusiones similares?

18. En la siguiente tabla están los datos de un diseño factorial 2 × 3 con tres réplicas, conteste:

		Nivel de *B*			
		B_1	B_2	B_3	Total $Y_{i\cdot\cdot}$
Niveles de *A*	A_1	44 34 $Y_{11\cdot}$ = 30	49 48 $Y_{12\cdot}$ = 43	52 58 $Y_{13\cdot}$ = 62	
	A_2	62 56 $Y_{21\cdot}$ = 58	68 70 $Y_{22\cdot}$ = 58	69 75 $Y_{23\cdot}$ = 65	
Total $Y_{\cdot j\cdot}$					$Y_{\cdots}$ =

a) Complete los totales que se piden en la tabla anterior.
b) Calcule las sumas de cuadrados correspondientes: SC_A, SC_B, SC_{AB}, SC_T y SC_E.
c) Obtenga la tabla de análisis de varianza y anote las principales conclusiones.
d) Sin tomar en cuenta el posible efecto de interacción, obtenga la diferencia mínima significativa (LSD) para comparar las medias en los niveles de factor *A* y la LSD para comparar las medias de *Y* en los niveles del factor *B*.
e) ¿Cuál sería la LSD exacta tomando en cuenta la interacción?

19. Se corre un diseño factorial 3 × 2 con 10 réplicas para investigar el hinchamiento del catalizador después de la extrusión en la fabricación de botellas de polietileno de alta densidad. El catalizador se utiliza en la obtención de dicho polietileno. Los factores investigados son: molde (con dos niveles) y *B*: catalizador (con tres niveles). Los datos obtenidos se muestran en la siguiente tabla:

		Catalizador								
		B_1			B_2			B_3		
Molde	A_1	93	92	93	92	90	92	95	94	94
		92	91	90	94	91	91	94	97	96
		90	90		90	92		94	95	
		91	91		91	92		94	96	
	A_2	88	88	87	90	89	88	91	97	91
		88	87	88	88	90	89	90	89	91
		87	87		88	89		92	90	
		87	87		88	88		90	91	

a) Plantee las hipótesis de interés en este problema y el modelo estadístico correspondiente.
b) Construya la tabla de análisis de varianza y determine cuáles efectos están activos.
c) Dibuje las gráficas de medias para los dos efectos principales con los métodos LSD y de Tukey. Compare los resultados de ambos métodos.
d) Haga la gráfica de interacción con intervalos de confianza sobrepuestos.
e) Determine cuál es el mejor tratamiento. ¿Cuál es el hinchamiento predicho en el mejor tratamiento?
f) Verifique los supuestos de normalidad y varianza constante.
g) Utilice la gráfica de residuos contra factores para detectar posibles efectos sobre la dispersión del hinchamiento. ¿En cuál molde parece que es menor la dispersión?

20. Para mejorar la resistencia a la torsión de las adhesiones de componentes electrónicos sobre placas, se estudiaron dos tipos de pegamentos (A_1 y A_2) y tres temperaturas de curado (60, 80 y 100°C). En cada combinación se analizaron dos componentes y los resultados obtenidos son los siguientes:

	Curado		
	60	80	100
Pegamento A_1	2.5	3.8	4.0
	2.8	3.4	4.2
Pegamento A_2	1.6	3.2	4.3
	1.22	2.8	4.7

a) Plantee las hipótesis de interés en este problema y el modelo estadístico correspondiente.
b) Construya el ANOVA y decida cuáles efectos están activos.
c) Dibuje las gráficas de efectos y determine con ellas el mejor tratamiento.
d) Estime la resistencia a la torsión en el mejor tratamiento.
e) Verifique residuos.

21. Se desea investigar de qué manera afecta el tiempo de curado y el tipo del acelerante a la resistencia de caucho vulcanizado. Se realiza un experimento y se obtienen los siguientes datos:

Tiempo de cura a 14°C (minutos)	Acelerante		
	A	*B*	*C*
40	3 900, 3 600	4 300, 3 700	3 700, 4 100
60	4 100, 3 500	4 200, 3 900	3 900, 4 000
80	4 000, 3 800	4 300, 3 600	3 600, 3 800

a) Señale el nombre del diseño de experimento utilizado y su modelo estadístico.
b) Formule claramente todas las hipótesis que se pueden probar.
c) Realice el análisis estadístico apropiado para probar las hipótesis que formuló.
d) ¿Hay algún tiempo de cura que es mejor para aumentar la resistencia? Argumente su respuesta.
e) ¿Algún acelerante es mejor? Explique.
f) ¿Hay alguna combinación de tiempo y acelerante que sea mejor?
g) Explique de manera gráfica cómo se obtuvo en la computadora el valor-*p* para tiempo de cura.
h) Verifique que se cumplan los supuestos. En caso de que no se cumpliera el supuesto de varianza constante para el tiempo de cura, ¿qué significaría eso y cómo pudiera corregirse?

22. En una fábrica de aceites vegetales comestibles la calidad resulta afectada por la cantidad de impurezas dentro del aceite, ya que éstas causan oxidación, y ello repercute a su vez en las características de sabor y color del producto final. El proceso de "blanqueo" es el responsable de eliminar tales impurezas, y una forma de medir su eficacia es midiendo el color del aceite. Para generar una primera aproximación a la solución del problema se decide estudiar el efecto de la temperatura y el porcentaje de arcilla en el color del aceite inicialmente a nivel laboratorio. El diseño y los datos de las pruebas experimentales se muestran a continuación.

Temperatura	Porcentaje de arcilla			
	0.8	0.9	1.0	1.1
90	5.8 5.9	5.4 5.5	4.9 5.1	4.5 4.4
100	5.0 4.9	4.8 4.7	4.6 4.4	4.1 4.3
110	4.7 4.6	4.4 4.4	4.1 4.0	3.7 3.6

a) Construya el modelo estadístico y formule las hipótesis pertinentes.
b) ¿Cuál es el nombre del diseño utilizado?

c) Por lo general, a condiciones reales se utiliza 1.1% de arcilla y 100 grados de temperatura. ¿Por qué cree que se eligieron precisamente esos niveles de prueba para el experimento?
d) Realice un análisis de varianza para probar las hipótesis y obtenga conclusiones.
e) Apoyándose en las gráficas de efectos, ¿cuál es la relación general entre el color y los factores controlados en su rango de experimentación?
f) A partir de la gráfica de interacciones, ¿cree que haya un efecto no lineal?
g) Considerando que el nivel mínimo aceptable de blancura es de 4.8, ¿qué tratamiento utilizaría?
h) ¿Vale la pena plantear el estudio en condiciones reales?
i) ¿Qué cambio le haría al experimento si lo corre en condiciones reales?

23. En un laboratorio de microbiología se realiza un experimento para investigar si influye el tipo de verdura (lechuga-L, cilantro-C, zanahoria-Z) y la temperatura (8 y 20°C) de almacenamiento en la sobrevivencia del vidrio colerae. Se hicieron varias réplicas. El porcentaje de sobrevivencia obtenido después de 24 horas de inoculado el alimento se muestra a continuación:

	Temperatura	
Alimento	**20**	**8**
L	13.1 15.0 33.6 35.5 42.0 11.1 12.8	6.2 28.5 41.0 35.9 25.0 23.8 79.0 41.6
C	19.0 19.0 66.6 66.6 11.0 11.0 49.0 49.0	84.3 68.7 68.7 30.5 30.5 11.0 11.0 20.0
Z	1.2 1.2 0.2 0.1 0.3 0.2 0.1 0.4 0.2 0.3	25.8 21.8 16.0 16.0 20.1 15.4 13.3 25.2

a) Señale el nombre del diseño empleado y formule las hipótesis que pueden ser probadas.
b) Haga un análisis de varianza e interprételo con detalle.
c) Verifique el supuesto de igual varianza entre los tratamientos (varianza constante). ¿Se cumple satisfactoriamente?
d) En caso de que no se cumpla el supuesto anterior, ¿cómo afecta esto a sus conclusiones?

24. Para el caso del problema anterior:
a) Transforme los datos con logaritmos y haga el análisis de varianza.
b) Verifique supuestos.
c) ¿Cuáles son las diferencias más importantes que encontró en los dos análisis? ¿Por qué?
d) Con los datos transformados, y en caso de que haya alguna interacción relevante, interprétela con detalle.

25. Con el objetivo de estudiar la producción de huitlacoche (hongo comestible del maíz) se decide correr un experimento con tres variedades de maíz en dos localidades maiceras. Las variables de interés fueron: el porcentaje de cobertura de la mazorca por el hongo, el peso total de la mazorca y el peso del huitlacoche. Se hicieron cuatro réplicas. Los datos obtenidos que representan promedios de 20 mazorcas infectadas se muestran en la siguiente tabla:

Variedad	Localidad	Cobertura	Peso *T*	Peso *H*
A	*A*	74.5	353.3	145.9
A	*A*	49.5	239.2	119.4
A	*A*	75.5	358.9	121.8
A	*A*	74.4	310.8	121.6
A	*B*	41.3	295.1	192.2
A	*B*	39.4	282.2	195.0
A	*B*	40.6	240.1	177.0
A	*B*	22.9	191.7	142.3
B	*A*	67.1	386.3	175.1
B	*A*	62.9	231.0	138.0
B	*A*	57.1	264.8	116.4
B	*A*	66.3	180.8	73.9
B	*B*	35.0	255.8	169.6
B	*B*	22.3	230.2	180.0
B	*B*	30.0	242.4	168.8
B	*B*	39.6	224.7	163.0
C	*A*	72.7	320.0	139.1
C	*A*	72.6	277.1	150.0
C	*A*	64.8	306.4	120.4
C	*A*	70.6	327.8	110.0
C	*B*	32.0	281.2	182.7
C	*B*	31.9	220.0	198.1
C	*B*	35.7	265.5	205.6
C	*B*	21.7	202.5	150.0

a) Escriba el nombre y modelo estadístico del diseño que se está empleando.
b) ¿Hay un efecto significativo de los factores variedad y localidad en las tres variables de respuesta?
c) ¿Existe claramente una localidad y variedad de maíz donde se produce más huitlacoche? Apóyese en gráficas y pruebas estadísticas.
d) ¿El que haya mayor cobertura del hongo garantiza mayor producción de huitlacoche? Vea de manera simultánea las gráficas de interacción para ambas variables de respuesta y/o haga un análisis de correlación entre estas dos variables.
e) ¿Cuánto huitlacoche se deja de producir en promedio en la localidad *A*?
f) Haga los análisis de residuos para verificar los supuestos del modelo.

26. Los siguientes datos corresponden a diseño 3 × 3 con tres réplicas. Interesa investigar el efecto de ambos factores sobre *Y*, para encontrar las condiciones adecuadas para maximizar.

	B		
A	B_1	B_2	B_3
A_1	10 6 14	3 5 1	1 2 1
A_2	60 73 79	88 70 76	71 71 69
A_3	44 35 28	38 22 26	29 20 22

a) Especifique el modelo estadístico para el problema y las hipótesis pertinentes.
b) Haga un análisis de varianza y obtenga conclusiones.
c) Interprete con detalle el efecto de interacción, si es significativo.
d) Verifique supuestos.
e) ¿Hay un tratamiento mejor? Argumente con pruebas estadísticas.

27. En una empresa alimenticia se desean evaluar cuatro antioxidantes, a través de su efecto en un aceite vegetal. El propósito es seleccionar el producto que retrase más la oxidación. Las pruebas se hacen en condiciones de estrés, midiendo como variable de respuesta al índice de peróxidos. Se evalúan diferentes unidades experimentales a diferentes tiempos. Los datos obtenidos se muestran a continuación (en el control no se agrega ningún antioxidante). Dado que uno de los factores es el tiempo, y éste no se puede aleatorizar, entonces se le puede ver como un factor de bloques.

Producto	Tiempo		
	4 horas	8 horas	12 horas
Control	3.84, 3.72	27.63, 27.58	39.95, 39.00
A	4.00, 3.91	22.00, 21.83	46.20, 45.60
B	3.61, 3.61	21.94, 21.85	46.58, 42.98
C	3.57, 3.50	20.50, 20.32	45.14, 44.89
D	3.64, 3.61	20.30, 20.19	44.36, 44.02

a) Señale los factores controlados y la variable de respuesta.
b) Formule el modelo estadístico más apropiado al problema y las hipótesis estadísticas que se pueden probar.
c) Haga un análisis de varianza y observe los aspectos más relevantes.
d) ¿Los supuestos del modelo se cumplen?
e) Considerando que a menor índice de peróxidos mejor es el producto, ¿hay algún producto que sea mejor estadísticamente?

28. Se cree que la adhesividad de un pegamento depende de la presión y de la temperatura al ser aplicado. Se realiza un experimento factorial con ambos factores fijos.

Presión (lb/pulg²)	Temperatura (°F)		
	250	260	270
120	9.60	11.28	9.00
130	9.69	10.10	9.57
140	8.43	11.01	9.03
150	9.98	10.44	9.80

a) Formule las hipótesis y el modelo estadístico que se desea probar.
b) Analice los datos y obtenga las conclusiones apropiadas.
c) ¿Se puede analizar si hay interacción entre los dos factores controlados?
d) Verifique residuos.

29. Vuelva a analizar los datos del ejemplo 5.4 considerando a los inspectores como fijos. Plantee las hipótesis de interés en este caso y pruébelas. Estime los componentes de varianza. Por último, saque conclusiones en el contexto del problema.

Capítulo 6
Diseños factoriales 2^k

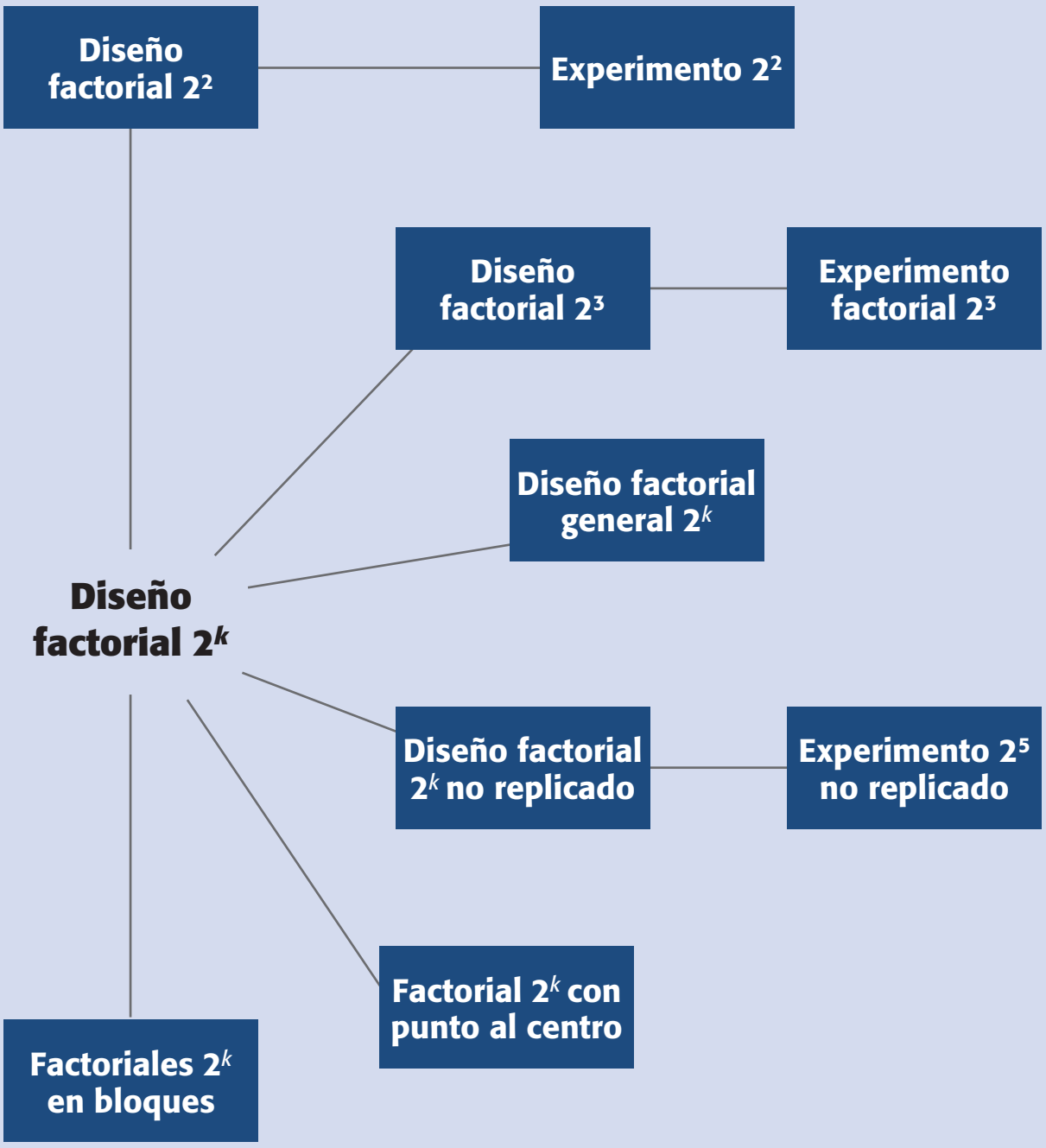
Sumario

- Diseño factorial 2^2
- Experimento 2^2: ejemplo integrador
- Diseño factorial 2^3
- Experimento 2^3: ejemplo integrador
- Diseño factorial general 2^k
- Diseño factorial 2^k no replicado
- Experimento 2^5 no replicado: ejemplo integrador
- Cuando la significancia de los efectos es menos clara: un ejemplo
- Factoriales 2^k con punto al centro
- Factoriales 2^k en bloques
- Uso de software estadístico

Objetivos de aprendizaje

- Conocer y aplicar los aspectos fundamentales de los diseños factoriales 2^k y tomar decisiones acerca de cuándo se debe aplicar cada diseño.
- Saber diseñar un experimento factorial 2^5 no replicado para aplicarlo a diversos casos.
- Identificar cuándo y por qué aplicar el diseño factorial 2^k en bloques o con punto al centro.

Mapa conceptual



Conceptos clave

- Coeficientes de determinación
- Columna de signos para contrastes
- Contraste
- Diagrama de Pareto de efectos
- Diseño factorial 2^2
- Diseño factorial 2^3
- Error estándar de un estadístico
- Factoriales 2^k completos
- Gráfica de contornos
- Gráfico de efectos en papel normal
- Mejor ANOVA
- Notación de Yates
- Papel probabilístico medio normal
- Proyectar el diseño
- Punto al centro
- R^2, R^2_{aj}
- Región experimental
- Residual
- Superficie de respuesta
- Tabla de signos
- Tamaño de prueba o corrida
- Unidades originales

Factoriales 2^k completos
Diseño que estudia k factores con 2 niveles cada uno, y donde se corren los 2^k posibles tratamientos. Son útiles cuando $2 \leq k \leq 5$.

Diseño factorial 2^2
Modelo que estudia el efecto de dos factores considerando dos niveles en cada uno.

En este capítulo se presenta la familia de diseños factoriales completos 2^k (k factores con dos niveles de prueba cada uno), que es una de las familias de diseños de mayor impacto en la industria y en la investigación, debido a su eficacia y versatilidad. Los *factoriales 2^k completos* son útiles principalmente cuando el número de factores a estudiar está entre dos y cinco ($2 \leq k \leq 5$), rango en el cual su tamaño se encuentra entre cuatro y 32 tratamientos; esta cantidad es manejable en muchas situaciones experimentales. Si el número de factores es mayor que cinco se recomienda utilizar un factorial fraccionado 2^{k-p} (capítulo 8). En general, los factoriales en dos niveles, sean completos o fraccionados, constituyen el conjunto de diseños de mayor impacto en las aplicaciones.

En el capítulo 5 se introdujo el concepto de *diseño factorial* y se definieron los efectos principales y efectos de interacción en el contexto de un arreglo factorial 2^2, que es el diseño más simple de la familia 2^k. Veamos con mayor detalle en qué consiste el factorial 2^2, su análisis e interpretación de resultados.

Diseño factorial 2^2

Con un *diseño factorial* 2^2 se estudia el efecto de dos factores considerando dos niveles en cada uno. Cada réplica de este diseño consiste de $2 \times 2 = 4$ combinaciones o tratamientos que se pueden denotar de diferentes maneras, como se muestra en la tabla 6.1. Algunas de estas notaciones se utilizan en situaciones muy particulares; por ejemplo, la notación +1, –1 es útil a la hora de hacer los cálculos para ajustar por mínimos cuadrados un modelo de regresión a los datos; es la notación que utilizan los paquetes *Statgraphics* y *Minitab*.

La notación de signos +, – es muy práctica para escribir las matrices de diseño; esta notación, combinada con la de Yates (véase última columna de la tabla 6.1) permite representar y calcular fácilmente los efectos de interés. La notación con letras A^+, A^- se utiliza para escribir, al final del análisis del experimento, el mejor punto o tratamiento ganador que se ha encontrado.

La notación de Yates [(1), a, b, ab] tiene un significado diferente a las demás: con ella se representa el *total* o la *suma de las observaciones en cada tratamiento*, más que al tratamiento mismo. Hay que observar que la lógica de la notación de Yates es la siguiente: si una letra minúscula está presente, entonces el factor correspondiente se encuentra en su nivel alto; si está ausente, el factor está en su nivel bajo; por ejemplo, *ab* se refiere al tratamiento en el que los factores A y B están en su nivel alto.

Tabla 6.1 Seis maneras de escribir los tratamientos del diseño 2^2.

	A	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*	Notación de Yates
Trat 1 →	bajo	bajo	A_1	B_1	A^-	B^-	–	–	0	0	–1	–1	(1)
Trat 2 →	alto	bajo	A_2	B_1	A^+	B^-	+	–	1	0	1	–1	*a*
Trat 3 →	bajo	alto	A_1	B_2	A^-	B^+	–	+	0	1	–1	1	*b*
Trat 4 →	alto	alto	A_2	B_2	A^+	B^+	+	+	1	1	1	1	*ab*

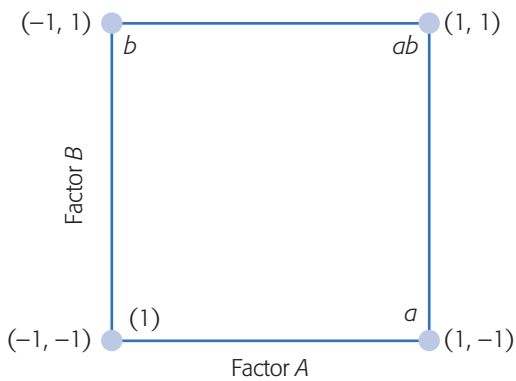


Figura 6.1 Representación del diseño factorial 2^2.

Representación geométrica

El diseño factorial 2^2 se representa de manera geométrica por los vértices del cuadrado de la figura 6.1. Cada vértice representa un punto de diseño o tratamiento. El área limitada por este cuadrado se conoce como *región experimental* y, en principio, las conclusiones que se obtengan del experimento sólo tienen validez sobre esta región.

Región experimental
Espacio delimitado por los rangos de experimentación utilizados con cada factor; las conclusiones del experimento son válidas principalmente en esta región.

Cálculo de los efectos

En este diseño hay tres efectos de interés: los dos efectos principales (A y B) y el efecto de interacción (AB). Con el uso de la notación de Yates podemos ver que si cada tratamiento se corre n veces, entonces la media de Y en el nivel alto de A es $(a + ab)/2n$ y en el nivel bajo es $(b + (1))/2n$. De aquí, y de la definición de efecto dada en el capítulo anterior, el *efecto A* se calcula como:

$$\text{Efecto } A = \frac{1}{2n}[a + ab - b - (1)] = \frac{[a + ab]}{2n} - \frac{[b + (1)]}{2n} \qquad \textbf{(6.1)}$$

y el *efecto B* es,

$$\text{Efecto } B = \frac{1}{2n}[b + ab - a - (1)] = \frac{[b + ab]}{2n} - \frac{[a + (1)]}{2n} \qquad \textbf{(6.2)}$$

En la parte derecha de estas expresiones se hace evidente que los efectos son diferencias de medias. En sentido geométrico, el efecto A equivale a promediar los datos del lado derecho del cuadrado de la figura 6.1 y restarles el promedio de los datos del lado izquierdo; mientras que para el *efecto B* se promedian los datos del lado de arriba y se le resta la media de los datos del lado de abajo.

El efecto de interacción entre los factores A y B está dado por la diferencia entre el efecto de A en el nivel alto de B y el efecto de A en el nivel bajo de B, esto es,

$$\text{Efecto } AB = \frac{1}{2n}[ab + (1) - a - b] = \frac{[ab - b]}{2n} - \frac{[a - (1)]}{2n} \qquad \textbf{(6.3)}$$

que también es una diferencia de medias. En términos geométricos, la interacción es la diferencia entre las medias de las diagonales del cuadrado de la figura 6.1. Si los datos del ejemplo 5.1 se retoman de la tabla 5.1, se tiene que $(1) = 28$, $a = 41$, $b = 63$,

$ab = 45$. Con las expresiones dadas por las ecuaciones (6.1), (6.2) y (6.3) los efectos resultan ser:

$$\text{efecto } A = -2.5, \text{ efecto } B = 19.5 \text{ y efecto } AB = -15.5 \tag{6.4}$$

que ya se habían obtenido en el ejemplo.

Análisis de varianza

Aunque los efectos calculados con datos muestrales sean números distintos de cero, esto no implica que el efecto correspondiente sea *estadísticamente* diferente de cero. O, si en su representación gráfica aparentan ser importantes, eso tampoco es suficiente para concluir que afectan de manera significativa la variable de respuesta. Para afirmar que tales efectos contribuyen a explicar el comportamiento de la respuesta, se debe hacer la prueba estadística del análisis de varianza. Las sumas de cuadrados que componen el ANOVA se pueden calcular como se indicó en el capítulo 5 o también por medio de los efectos estimados, como veremos a continuación. Es obvio que si se tiene un software especializado que haga los cálculos, lo que sigue sirve de fundamento.

Definición de contraste

Contraste
Es una combinación lineal de la forma $C = \Sigma_{i=1}^{2^k} c_i Y_i$, con $\Sigma c_i = 0$

Una combinación lineal de la forma $C = \Sigma_{i=1}^{2^k} c_i Y_i$, con $\Sigma c_i = 0$ se llama *contraste*. Se sabe que la suma de cuadrados para cualquier contraste C está dada por

$$SC_C = \frac{\left(\sum_{i=1}^{2^k} c_i Y_i\right)^2}{n\sum_{i=1}^{2^k} c_i^2} \tag{6.5}$$

la cual sólo tiene un grado de libertad. Note que en el contexto de los diseños factoriales, las sumas corren sobre los 2^k tratamientos del diseño factorial 2^k, y cada término Y_i representa un término de la notación de Yates. Por ejemplo, los contrastes correspondientes a los tres efectos A, B y AB en el diseño factorial 2^2 están dados por:

$$\text{Contraste } A = [a + ab - b - (1)],$$

$$\text{Contraste } B = [b + ab - a - (1)]$$

$$\text{Contraste } AB = [ab + (1) - a - b]$$

que, como ya hemos visto, son las cantidades que definen a los efectos. Son contrastes por el hecho de que son combinaciones lineales donde los coeficientes suman cero ($1 + 1 - 1 - 1 = 0$). Una vez obtenido el contraste, el efecto correspondiente se obtiene dividiéndolo entre la constante que lo convierta en una diferencia de medias; este número es la mitad de las observaciones hechas en el experimento [véanse ecuaciones (6.1), (6.2) y (6.3)]. Por ejemplo, en el factoria1 2^k con n réplicas los contrastes se dividen por $n2^{(k-l)}$ para estimar los efectos; en particular para el diseño 2^2 con n réplicas se divide por $n2^{(2-1)} = 2n$.

Métodos para calcular contrastes. Una manera práctica de establecer los contrastes de cualquier efecto, sea principal o de interacción en el diseño factorial 2^k es mediante el auxilio de una tabla de signos. La *tabla de signos* se construye a partir de la matriz de diseño, multiplicando las columnas que intervienen en la interacción que se quiera calcular. Por ejemplo, si se quiere obtener el contraste de la interacción doble *AB*, se multiplica la columna de signos *A* por la columna *B*, y el resultado son los signos de contraste *AB*. Esto se muestra en la siguiente tabla de signos para el diseño factorial 2^2.

Tabla de signos
Es una manera práctica de obtener el contraste de cualquier efecto. Se construye a partir de la matriz de diseño, multiplicando las columnas que intervienen en la interacción calculada.

A	*B*	*AB*	Yates
−	−	+	(1)
+	−	−	*a*
−	+	−	*b*
+	+	+	*ab*

En la tabla de signos, las columnas que corresponden a los efectos principales coinciden con la matriz de diseño. Una vez obtenidas las columnas de signos de los efectos de interés, el contraste de cada efecto resulta de multiplicar su columna de signos por la columna de los datos expresados en la notación de Yates. Recordemos que la *notación de Yates* representa los totales o sumas de las observaciones en cada tratamiento. Por ejemplo, al multiplicar las columnas *A* y *B* por la notación de Yates, se obtiene el *contraste de AB* que ya conocemos: *Contraste AB* = $[(1) + ab - a - b]$. El cálculo de contrastes se complica en la medida de que el número de factores crece, dado que se incrementa rápidamente el número de términos que contiene el contraste. Por ejemplo, en un factorial 2^5 los contrastes están formados por 32 términos. De manera que a medida que se tengan más factores será necesario utilizar un software estadístico para hacer los cálculos.

Notación de Yates
Representa los totales o sumas de las observaciones en cada tratamiento de un diseño 2^k.

Pasos para llegar al ANOVA. Para obtener el ANOVA se necesita calcular la suma de cuadrados de cada uno de los efectos. Como se ilustra en la figura 6.2, las

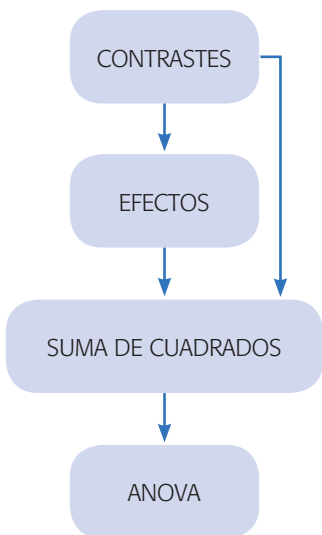


Figura 6.2 Pasos para llegar al ANOVA en un diseño 2^k.

sumas de cuadrados se pueden obtener ya sea de los efectos o directamente de los contrastes.

Una vez obtenidos los contrastes se procede a estimar los efectos, dividiendo éstos por la constante que los convierte en diferencias de medias. En particular, el diseño factorial 2^2 que está constante es $2n$. Para investigar cuáles de los tres efectos están activos o son significativos se procede a probar las hipótesis dadas por:

$$H_0 : \text{Efecto } A = 0$$

$$H_0 : \text{Efecto } B = 0$$

$$H_0 : \text{Efecto } AB = 0$$

cada una contra la alternativa de que el efecto en cuestión es diferente de cero. Estas hipótesis se prueban con el análisis de varianza.

Para obtener las sumas de cuadrados para cada efecto se aplica el resultado sobre contrastes de la ecuación 6.5, y resultan las expresiones dadas por,

$$SC_A = \frac{[a + ab - b - (1)]^2}{n2^2} \tag{6.6}$$

$$SC_B = \frac{[b + ab - a - (1)]^2}{n2^2} \tag{6.7}$$

$$SC_{AB} = \frac{[ab + (1) - a - b]^2}{n2^2} \tag{6.8}$$

donde cada una de ellas tiene sólo un grado de libertad, debido a que cada factor tiene únicamente dos niveles. La suma de cuadrados totales se calcula con la expresión,

$$SC_T = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{l=1}^{n} Y_{ijl}^2 - \frac{Y_{\cdots}^2}{n2^2}$$

y tiene $n2^{2-1}$ grados de libertad, es decir, el total de observaciones en el experimento menos uno. La suma de cuadrados del error se calcula por diferencia: $SC_E = SC_T - SC_A - SC_B - SC_{AB}$ y tiene $(n2^{2-1}) - 3 = 4(n-1)$ grados de libertad. La tabla de ANOVA del diseño factorial 2^2 con n réplicas se presenta en la tabla 6.2.

Recordemos que si el valor-p es menor que el nivel de significancia prefijado α, se concluye que el efecto correspondiente es estadísticamente distinto de cero, es decir, tal efecto está activo o influye de manera significativa sobre la respuesta. Además, mientras más pequeño sea el valor-p de un efecto, este último es más importante. En la tabla de ANOVA nótese que para calcular el CM_E se requieren al menos dos réplicas del experimento: con una réplica habría cero grados de libertad para el error, no se podría calcular su cuadrado medio y, por ende, no habría ANOVA. Como se verá más adelante, se recomienda correr el factorial 2^2 con al menos tres réplicas para poder estimar un CM_E confiable.

Interpretación de los efectos significativos. Con el análisis de varianza de la tabla 6.2 se sabe cuáles de los tres efectos A, B o AB actúan sobre la variable de

Tabla 6.2 ANOVA para el diseño factorial 2^2.

FV	SC	GL	CM	F_0	Valor-*p*
A	SC_A	1	CM_A	CM_A/CM_E	$P(F > F_0)$
B	SC_B	1	CM_B	CM_B/CM_E	$P(F > F_0)$
AB	SC_{AB}	1	CM_{AB}	CM_{AB}/CM_E	$P(F > F_0)$
Error	SC_E	$4(n-1)$	CM_E		
Total	SC_T	$n2^2-1$			

respuesta. Se recomienda graficar los efectos que fueron activos y aquellos sobre los que se tenía algún tipo de conjetura muy definida; a continuación, se hace una interpretación a detalle y ésta se convierte en conocimiento sobre el proceso o fenómeno investigado. Para determinar las mejores condiciones de operación del proceso se deben utilizar sólo los efectos que resultan estar activos en el ANOVA.

Experimento 2^2: ejemplo integrador

Ahora interesa estudiar el efecto del *tamaño de broca* (factor A) y de la *velocidad* (factor B) sobre la *vibración* de la ranuradora (respuesta Y). Para ello, se decide utilizar un diseño factorial 2^2 con cuatro réplicas, lo cual da un total de $4 \times 2^2 = 16$ corridas del proceso, que se realizan en orden aleatorio. El tamaño de la broca se prueba en 1/16 y en 1/8 de pulgada y la velocidad en 40 y 90 revoluciones por segundo, según se describe en la siguiente tabla:

Factor	Niveles		Unidad
A: Broca	1/16	1/8	pulg.
B: Velocidad	40	90	rps

Unidades originales
Es la forma real, no codificada, con la que se expresan o identifican los niveles de prueba de cada factor.

En la tabla 6.3 se muestra el diseño factorial utilizado en sus *unidades originales*, que son las que se necesitan al momento de hacer las pruebas o corridas del proceso. También se muestra la notación (+, –), y los datos obtenidos en las 16 pruebas.

En la última columna se muestra el total por tratamiento utilizando la notación de Yates.

La aplicación en este ejemplo de los principios básicos del diseño de experimentos se aprecia en el hecho de que se *aleatorizaron* las 16 corridas del proceso y se hicieron cuatro *repeticiones* de cada tratamiento; además, se supone que todo el

Tabla 6.3 Diseño y datos para ranuradora.

A: Broca	B: Veloc.	A	B	x_1	x_2	Vibración				Total
1/16	40	–	–	–1	–1	18.2	18.9	12.9	14.4	64.4 = (1)
1/8	40	+	–	+1	–1	27.2	24.0	22.4	22.5	96.1 = *a*
1/16	90	–	+	–1	+1	15.9	14.5	15.1	14.2	59.7 = *b*
1/8	90	+	+	+1	+1	41.0	43.9	36.3	39.9	161.1 = *ab*

experimento se corre en igualdad de circunstancias con respecto al resto de factores no estudiados (*principio de bloqueo*), es decir, los factores controlables no incluidos en el estudio se mantienen lo más fijo posible durante la realización de las pruebas.

La representación geométrica del experimento se muestra en la figura 6.3. Observe la relación entre las *unidades originales* y las *unidades codificadas*, así como el significado de la notación de Yates.

Las preguntas principales que se quieren responder con el experimento son: ¿la velocidad y el tamaño de la broca afectan la vibración de la ranuradora? Si la afectan, ¿cómo es tal efecto y cuál combinación de velocidad y tamaño de broca minimiza la vibración?, ¿cuál es la vibración esperada en las condiciones óptimas?, ¿se cumplen los supuestos del modelo?

Efectos estimados. De acuerdo con las relaciones (6.1), (6.2) y (6.3), y con la última columna de la tabla 6.4, los efectos estimados están dados por:

$$A = \frac{1}{2n}[a + ab - b - (1)] = \frac{2}{2(4)}[96.1 + 161.1 - 59.7 - 64.4] = 16.64 \qquad \textbf{(6.9)}$$

$$B = \frac{1}{2n}[b + ab - a - (1)] = \frac{1}{2(4)}[59.7 + 161.1 - 96.1 - 64.4] = 7.54 \qquad \textbf{(6.10)}$$

$$AB = \frac{1}{2n}[ab + (1) - a - b] = \frac{1}{2(4)}[161.1 + 64.4 - 96.1 - 59.7] = 8.71 \qquad \textbf{(6.11)}$$

Se observa que el efecto del *tamaño de broca* (factor A) es prácticamente el doble de los otros dos, pero falta investigar si alguno es estadísticamente significativo.

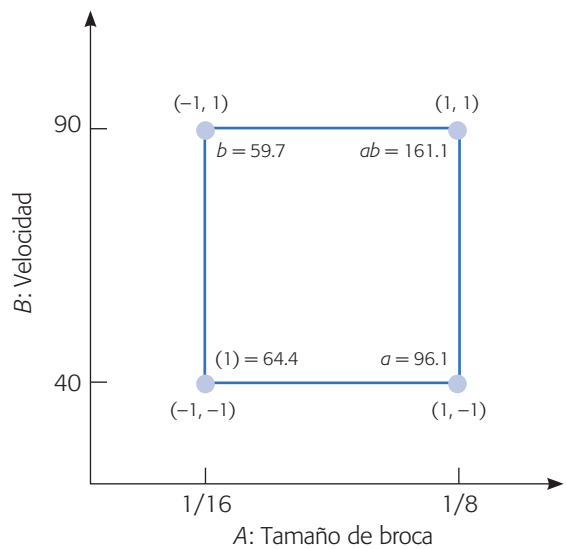


Figura 6.3 Representación geométrica.

Análisis de varianza. Las sumas de cuadrados (SC) de los efectos se calcula a partir de sus contrastes (ecuaciones 6.6, 6.7 y 6.8) como,

$$SC_A = \frac{[96.1 + 161.1 - 59.7 - 64.4]^2}{16} = 1\,107.22$$

$$SB_B = \frac{[59.7 + 161.1 - 96.1 - 64.4]^2}{16} = 227.25$$

$$SC_{AB} = \frac{[161.1 + 64.4 - 96.1 - 59.7]^2}{16} = 303.63$$

La suma de cuadrados totales es:

$$SC_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{l=1}^{n} Y_{ijl}^2 - \frac{Y_{\cdots}^2}{4n} = 1\,709.83$$

y tiene 15 grados de libertad, mientras que la suma de cuadrados del error se calcula por diferencia como,

$$SC_E = 1\,709.83 - 1\,107.22 - 227.25 - 303.63 = 71.73$$

y le quedan 15 – 3 = 12 grados de libertad. El análisis de varianza se muestra en la tabla 6.4. De acuerdo con la columna para el valor-p, cuyas entradas son menores que 0.05, se rechazan las tres hipótesis nulas H_0: *efecto* $A = 0$, H_0: *efecto* $B = 0$ y H_0: *efecto* $AB = 0$, y se concluye que H_A: *efecto* $A \neq 0$, H_B: *efecto* $B \neq 0$ y H_{AB}: *efecto* $AB \neq 0$, respectivamente, con $\alpha = 0.05$. Esto es, se acepta que *sí* hay efecto de A, B y AB, es decir, los tres efectos están activos o son significativos. El valor-p de magnitud tan pequeña para los tres efectos nos muestra que la conclusión es contundente y que prácticamente no se corre ningún riesgo de rechazar en falso. Del valor de F_0 se aprecia que el efecto más importante es el del factor A.

Interpretación y conclusiones. Antes de comenzar a interpretar las gráficas de los efectos que hayan resultado significativos en el análisis de varianza, debemos tener presente lo siguiente: que el objetivo es minimizar la vibración de la ranuradora y que las interacciones tienen prioridad con respecto a los efectos principales. En este caso, y de acuerdo con la ANOVA, tanto los dos efectos principales (*broca y velocidad*) como el efecto de interacción (*broca* × *velocidad*) tienen un efecto significativo sobre la vibración de la ranuradora. Puesto que sólo se estudian dos factores, toda la información relevante del experimento se encuentra en la gráfica de la interacción (figura 6.4 *a*). De esta figura se observa algo de la física del proceso: se puede afirmar

Tabla 6.4 ANOVA para el experimento de la ranuradora.

FV	SC	GL	CM	F_0	Valor-p
A: Broca	1 107.22	1	1 107.22	185.25	0.0000
B: Velocidad	227.25	1	227.25	38.02	0.0000
AB	303.63	1	303.63	50.80	0.0000
Error	71.73	12	5.98		
Total	1 709.83	15			

que cuando la broca se encuentra en su nivel bajo, la velocidad no afecta de manera significativa la vibración, por el contrario, cuando la broca se encuentra en su nivel alto, la velocidad tiene un efecto considerable sobre la vibración. En otras palabras, al estar la broca en su nivel bajo, la vibración será baja sin importar la velocidad. Si bien, es cierto que es razonable pensar que a mayor velocidad y a menor tamaño de broca la vibración sea menor, ésta es una conclusión *a posteriori*. Después de que ocurre siempre es posible encontrar una explicación razonable del hecho, pero lo interesante sería decirlo antes de observar el fenómeno y luego corroborar que en realidad se pensaba lo correcto. Pero, si lo que se quiere es minimizar la vibración, entonces se puede utilizar el tratamiento (A^-, B^+) o el (A^-, B^-), ya que ambos logran prácticamente los mismos resultados, por lo que la decisión de cuál de los dos utilizar se puede hacer con otros criterios, por ejemplo el tiempo de ciclo o el tiempo de vida del equipo. Si por alguna razón se tuviera que trabajar con la broca en nivel alto, entonces se debe trabajar a velocidad baja para que no se incremente tanto la vibración.

Con la interpretación hecha de la gráfica de interacción ya se tiene mucho conocimiento sobre el proceso, cosa que no se hubiese logrado si sólo nos hubiéramos

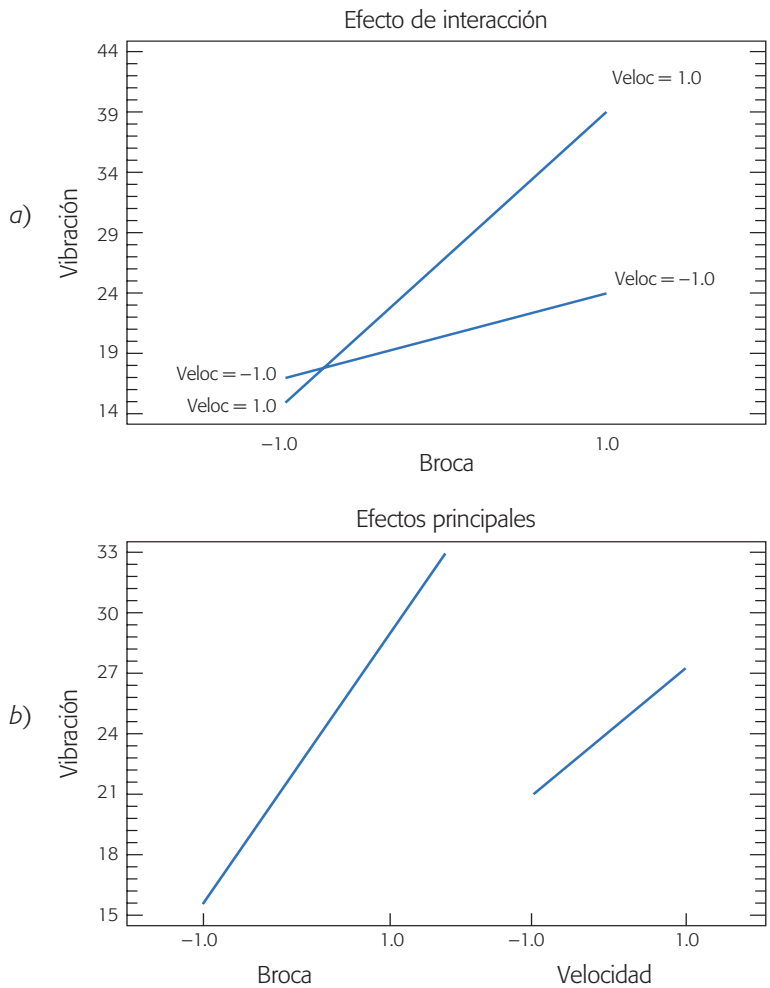


Figura 6.4 *a*) Efecto de interacción para ejemplo de la ranuradora. *b*) Gráfica de efectos principales para el ejemplo de la ranuradora.

limitado a interpretar los efectos principales de la figura 6.4*b*), de la cual se desprende que no se debe aumentar la velocidad, ni el tamaño de broca y que, por lo tanto, si se quiere minimizar la vibración forzosamente se debe utilizar la combinación (A^-, B^-). Es evidente que esta conclusión es más pobre que las que se obtuvieron del análisis de la gráfica de interacción. Cuando hay interacción, las conclusiones que se obtienen a partir de los efectos principales no siempre son ciertas. En general sólo se interpretan los efectos principales de aquellos factores que no interactúan con ningún otro.

La comparación de lo que se logró entender con la interacción permite destacar una vez más la importancia de ésta en los procesos, y recordar que con experimentación de prueba y error o moviendo un factor a la vez, prácticamente nunca se logrará captar el efecto de interacción. Por el contrario, el diseño estadístico de experimentos permite detectar y entender los efectos de interacción cuando los hay.

Modelo de regresión. Es útil ajustar un modelo de regresión (ver capítulo 11) a los datos experimentales con la finalidad de predecir el valor de Y en diferentes valores de los factores estudiados. Por ejemplo, en el caso del problema de la ranuradora, con el uso de valores codificados para los dos factores (x_1 y x_2), como se muestra en la tabla 6.3, el modelo de regresión ajustado que describe el comportamiento de la vibración sobre cualquier punto está dado por:

$$\hat{Y} = 23.83 + 8.32x_1 + 3.77x_2 + 4.35x_1x_2 \qquad \textbf{(6.12)}$$

donde $\hat{Y}$ es la respuesta predicha en el punto (x_1 y x_2), con $x_1 = A$: *broca* y $x_2 = B$: *velocidad.* En el caso de diseños 2^k, los coeficientes del modelo de regresión son iguales a los efectos estimados que resultaron significativos divididos entre dos. Así, por ejemplo, de acuerdo con (6.9), el coeficiente de x_1 es igual al efecto de $A/2$ = (16.64/2). Esta división entre dos se hace para lograr una escala unitaria, que es la escala usual en regresión. Los efectos originales no se encuentran en una escala unitaria, dado que el ancho de la región experimental es de dos unidades codificadas. El término independiente $\hat{\beta} = 23.83$ es la media global $\bar{Y}_{...}$ de todos los datos y representa la vibración predicha en el centro de la región experimental ($x_1 = 0$, $x_2 = 0$) (véase figura 6.5).

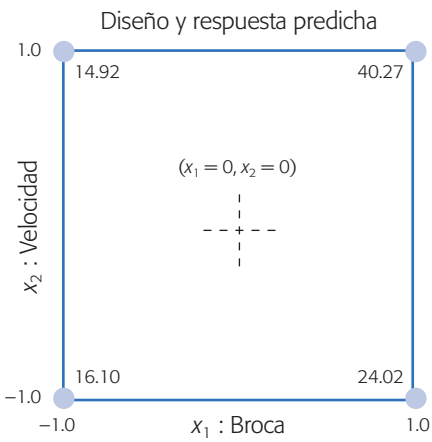


Figura 6.5 Región experimental y respuesta predicha para ranuradora.

La predicción del comportamiento de la vibración sobre la combinación de niveles que se quiera dentro de la región experimental se puede realizar con el modelo ajustado dado por la ecuación (6.12), evaluando el punto en la ecuación. La precisión de la predicción obtenida depende de la calidad del ajuste del modelo. A continuación vemos cómo medir la calidad del ajuste.

Coeficientes de determinación (R^2, R^2_{aj})
Miden la proporción o porcentaje de variabilidad en los datos experimentales que es explicada por el modelo considerado.

Coeficientes de determinación, R^2 y R^2_{aj}. Dos de los estadísticos más útiles para medir la calidad global del modelo de regresión múltiple es el *coeficiente de determinación* (R^2) y el *coeficiente de determinación ajustado* (R^2_{aj}), que se obtienen a partir del ANOVA de la siguiente manera:

$$R^2 = \frac{SC_{\text{total}} - SC_{\text{error}}}{SC_{\text{total}}} \times 100 = \frac{SC_{\text{modelo}}}{SC_{\text{total}}} \times 100$$

$$R^2_{aj} = \frac{CM_{\text{total}} - CM_{\text{error}}}{CM_{\text{total}}} \times 100$$

Nótese que estos coeficientes comparan la variabilidad explicada por el modelo frente a la variación total, cuantificadas a través de la suma de cuadrados (SC) o por el cuadrado medio (CM). De esta forma, para interpretar estos coeficientes se cumple que $0.0 \leq R^2_{aj} \leq R^2 \leq 100.0$ y cuantifican el porcentaje de variabilidad presente en los datos y que es explicado por el modelo; por ello, son deseables valores próximos a 100. En general, para fines de predicción se recomienda un coeficiente de determinación ajustado de al menos 70%. Cuando hay muchos factores se prefiere el estadístico R^2_{aj} en lugar del R^2, puesto que este último se incrementa de manera artificial con cada término que se agrega al modelo, aunque sea un término que no contribuya en mucho a la explicación de la respuesta. En cambio, el R^2_{aj} incluso baja de valor cuando el término que se agrega no aporta mucho.

Para el modelo de la vibración (experimento de la ranuradora) es directo calcular estos coeficientes a partir de la tabla 6.4:

$$R^2 = \frac{SC_{\text{total}} - SC_{\text{error}}}{SC_{\text{total}}} \times 100 = \frac{1\,709.83 - 71.73}{1\,709.83} \times 100 = 95.8$$

$$R^2_{aj} = \frac{CM_{\text{total}} - CM_{\text{error}}}{CM_{\text{total}}} \times 100 = \frac{\frac{1\,709.83}{15} - \frac{71.73}{12}}{\frac{1\,709.83}{15}} \times 100 = 94.76$$

De esta manera, de acuerdo con R^2_{aj}, el modelo ajustado de la expresión (6.12) explica 94.76% de la variabilidad de la vibración observada en el experimento (algo similar nos dice el R^2). Esto significa que los factores estudiados (tipo de broca y velocidad de la ranuradora), junto con su interacción, son responsables o explican un alto porcentaje de la variabilidad observada en la variable de respuesta (vibración). De esta manera, el efecto atribuible a factores no estudiados, ya sea que se hayan mantenido en un nivel fijo o que hayan tenido pequeñas variaciones, más el efecto de errores experimentales, fueron pequeños en comparación con el efecto de los factores estudiados.

En caso de que R^2_{aj} o R^2 hayan sido pequeños, esto indicaría que el efecto o variabilidad atribuible a los factores estudiados es pequeña en comparación con el

resto de la variación observada en el experimento. Un R^2 bajo puede deberse a una o varias de las siguientes razones:

- Los factores estudiados, por sí solos no tienen la suficiente influencia para explicar las variaciones observadas en la variable de respuesta.
- Los niveles de los factores estudiados son demasiado estrechos, por lo que el efecto sobre la variable de respuesta al cambiar de un nivel a otro es demasiado pequeño.
- Otros factores no estudiados en el experimento no se mantuvieron suficientemente fijos, por lo que al tener variaciones durante el experimento causaron mucha variación experimental.
- Los errores experimentales y los errores de medición fueron altos.

Se debe tener la precaución de analizar cuáles de las razones anteriores influyeron para tener coeficientes de determinación bajos; es decir, no se debe caer en el error de desechar el experimento y creer que "no sirvió". En general, los experimentos generan información que debe utilizarse para plantear conjeturas y estudios experimentales nuevos.

Hacer la predicción. La respuesta predicha $\hat{Y}$ en un punto dado es un estimador de la respuesta promedio en dicho punto. Por ejemplo, la predicción en uno de los mejores tratamientos (–1, 1) para minimizar se obtiene al sustituir este punto en el modelo ajustado (ecuación 6.12).

$$\hat{Y}(-1, 1) = 23.83 + 8.32(-1) + 3.77(1) + 4.35(-1)(1) = 14.92$$

De la misma manera, es posible sustituir cualquier punto de la región experimental en el modelo (6.12) y obtener la respuesta predicha sobre el punto; dicho valor es un estimador de la media de la vibración en ese tratamiento. En la figura 6.5 se presenta la región experimental y el valor de la respuesta predicha en cada punto de diseño. Por ejemplo, el peor tratamiento para minimizar es el punto (1, 1), donde se predice una vibración promedio de 40.27.

También se puede obtener un intervalo al 95% de confianza para la vibración promedio en uno de los mejores tratamientos (–1, 1), que está dado por [12.26, 17.58]. Se recomienda que este intervalo se obtenga con apoyo de un software. Los detalles sobre cómo obtener este intervalo de confianza se ven en el capítulo 11.

Gráficos de superficie. Existen dos gráficos de superficie que permiten tener una visualización de lo que significa el modelo ajustado dado en la ecuación 6.12. Éste representa una *superficie de respuesta* sobre la región experimental, que modela el comportamiento de la vibración. Estos gráficos del modelo son particularmente útiles cuando se tienen dos factores; sin embargo, para más de dos factores disminuyen su efectividad porque no se puede dibujar una superficie en cuatro o más dimensiones.

Superficie de respuesta
Es la superficie que resulta de representar gráficamente el modelo ajustado, y describe el comportamiento de la respuesta promedio en cada punto de la región experimental.

En la figura 6.6 se presenta el *gráfico de superficie*, el cual describe el comportamiento de la vibración sobre la región experimental. Las *curvas de nivel* o *isolíneas*, dibujadas como líneas más gruesas sobre la región experimental, son otra

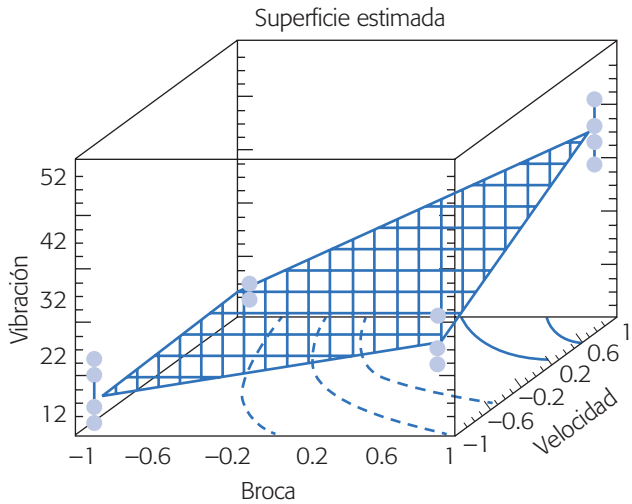


Figura 6.6 Superficie de respuesta del modelo ajustado en el experimento de la ranuradora.

manera de representar la superficie. Cada curva de nivel representa puntos o combinaciones de broca y velocidad donde la vibración es la misma.

Asimismo, se observa que los puntos donde la superficie toma valores más pequeños son precisamente en el mejor tratamiento que habíamos encontrado: (*broca baja*, *velocidad alta*) y (*broca baja*, *velocidad baja*). Se ve que la clave de la vibración pequeña es la broca en su nivel bajo, que es donde la superficie toma su menor altura. Los puntos en cada esquina de la superficie representan los datos del experimento. Note que la superficie trata de ajustarse lo mejor posible a los puntos observados.

Gráfica de contornos
Superficie de respuesta con curvas de nivel o isolíneas que permite ubicar los niveles de los factores sobre los cuales la variable de interés toma el mismo valor.

Para una mejor visualización, la representación en curvas de nivel o *gráfico de contornos* se puede dibujar sin la superficie, como se muestra en la figura 6.7. Esta representación es mejor que la de superficie porque se pueden ver con bastante exactitud las coordenadas del punto con la vibración deseada. Los números que acompañan a las curvas de nivel son precisamente la altura de la superficie sobre toda la *isolínea*, y éstas se pueden dibujar con los valores que el experimentador quiera.

Por ejemplo, como la curva con altitud igual a 16.0 pasa casi sobre el punto (–1, –1), esto implica que el modelo estimado evaluado en esta combinación debe predecir un valor de vibración muy cercano a 16.0. Al hacer los cálculos "exactos" con el modelo se observa que efectivamente $\hat{Y}(-1, -1) = 16.1$.

Residuos
Es la diferencia entre el valor observado en cierto tratamiento y la respuesta predicha por el modelo para tal tratamiento.

Análisis de residuos. El *residuo* o *residual* se define como la diferencia entre el valor observado en cierto tratamiento y la respuesta predicha por el modelo para tal tratamiento. Como se ha visto desde el capítulo 3, los residuos permiten evaluar varios aspectos de la calidad del modelo (sea un modelo de análisis de varianza o de regresión) que se propone para los datos, ya que en la medida de que los residuos sean pequeños, el modelo describirá de mejor manera el comportamiento de la respuesta.

Ahora, retomando la gráfica del modelo ajustado representada en la figura 6.6, en cada esquina de la superficie se observan las cuatro mediciones de vibración

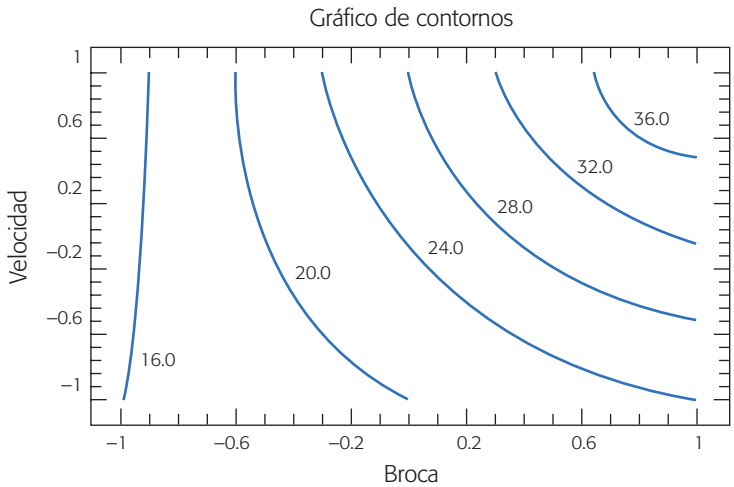


Figura 6.7 Gráfica de contornos de la vibración predicha, experimento de la ranuradora.

hechas en cada tratamiento. Si bien, la superficie trata de pasar lo más cerca posible de todos los puntos, no lo logra y por cada observación da lugar a un error llamado *residuo*. Siempre se tienen tantos residuos como datos y la media de los residuos es igual a cero. En general, el residuo correspondiente al dato Y_{ijl} está dado por

$$e_{ijl} = \text{observado} - \text{predicho} = Y_{ijk} - \hat{Y}_{ijk}$$

En la figura 6.8 se muestran los 16 residuos que corresponden a los 16 datos del experimento y se da la desviación estándar S de los residuos en cada tratamiento. Por ejemplo, el residuo $e_{121} = 0.975$ es la diferencia entre la primera medición en el punto (–, +) y el valor predicho por el modelo para este tratamiento, es decir,

$$e_{121} = Y_{121} - \hat{Y}_{121} = 15.9 - 14.925 = 0.975$$

La desviación estándar de los residuos en cada combinación indica el tratamiento que, de manera muestral, generó menor variabilidad en Y, y en este caso uno de los dos mejores tratamientos (–, +) en media, también es el que tiene menor variabilidad muestral ($S(-, +) = 0.75$). Una prueba estadística para probar la hipótesis de igualdad de varianzas en dos tratamientos diferentes ($H_0 : \sigma^2_{(i,j)} = \sigma^2_{(i,j)}$, con $(i, j) \neq (i, j)'$), se basa en el estadístico de prueba dado por:

$$Z_0^* = \ln\left[\frac{S^2_{(i,j)}}{S^2_{(i,j)'}}\right]$$

el cual se compara con los cuantiles de una distribución normal estándar. Se rechaza H_0 si $|Z_0^*|$ es mayor que Z_α. Por ejemplo, si se comparan las varianzas en las casillas (–, –) y (–, +), el estadístico calculado es $Z_0^* = \ln(8.46/0.56) = 2.71$. Como es mayor que 1.96, se concluye que las varianzas en esas combinaciones son estadísticamente diferentes. Esto es, la velocidad tiene efecto sobre la dispersión en el nivel bajo de la broca.

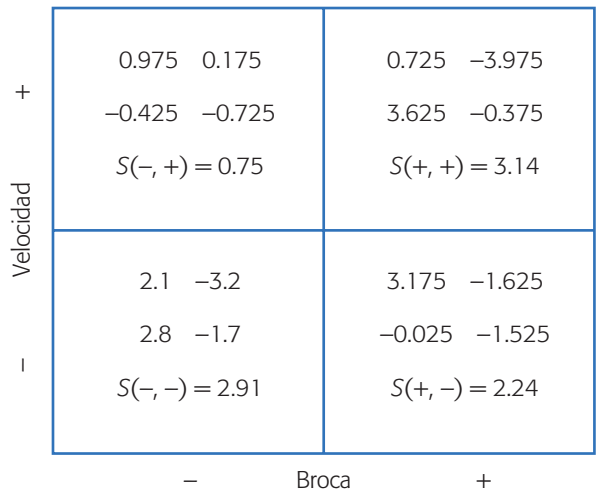


Figura 6.8 Valor de los residuos, experimento de la ranuradora.

Verificación de supuestos. Los supuestos del modelo de análisis de varianza deben verificarse antes de dar por válidas las conclusiones de la etapa de interpretación. Como se vio en el capítulo 3, la tabla de ANOVA (6.4) supone que los residuos se distribuyen *normales*, *independientes* y con *varianza constante*. La violación grave de cualquiera de estos supuestos conduce a conclusiones erróneas.

El supuesto de varianza constante se puede verificar graficando los residuos contra los predichos, y los puntos deben caer aleatoriamente en el sentido vertical dentro de una banda horizontal, para concluir que el supuesto se cumple (véase figura 6.9*a*). Los huecos en el sentido horizontal se deben a que sólo son cuatro puntos de diseño, y las predicciones en ellos difieren bastante. Aquí habría cierta evidencia para decir que el supuesto de varianza constante no se cumple, dada la dispersión más compacta de la primera columna de puntos en relación con las otras tres. Sin embargo, en este caso no es una situación grave que pudiera afectar las conclusiones del ANOVA, basadas en valores-*p* claramente pequeños. Además, la menor dispersión ocurre justo en el punto donde la vibración es menor, lo cual refuerza las conclusiones obtenidas. Otro gráfico que sirve para comprobar el supuesto de varianza constante es el de residuos contra cada factor. En la figura 6.9*b* se dibujan los residuos contra el factor broca. Se observa que las columnas de puntos en los niveles bajo y alto del factor broca, tienen "más o menos" la misma dispersión, por lo tanto se cumple el supuesto de varianza constante.

En la figura 6.9*c* se grafican los residuos en papel probabilístico normal. Como los puntos se apegan a la línea colocada visualmente (no es línea de regresión), se concluye que no hay violaciones al supuesto de normalidad.

Para probar el supuesto de independencia se requiere capturar los datos en el orden en que fueron obtenidos, a fin de que el software pueda graficar los residuos con respecto al tiempo u orden de corrida (figura 6.9*d*). Al no observarse ninguna tendencia en los puntos (que suban o bajen) se concluye que no existe problema con la declaración de independencia.

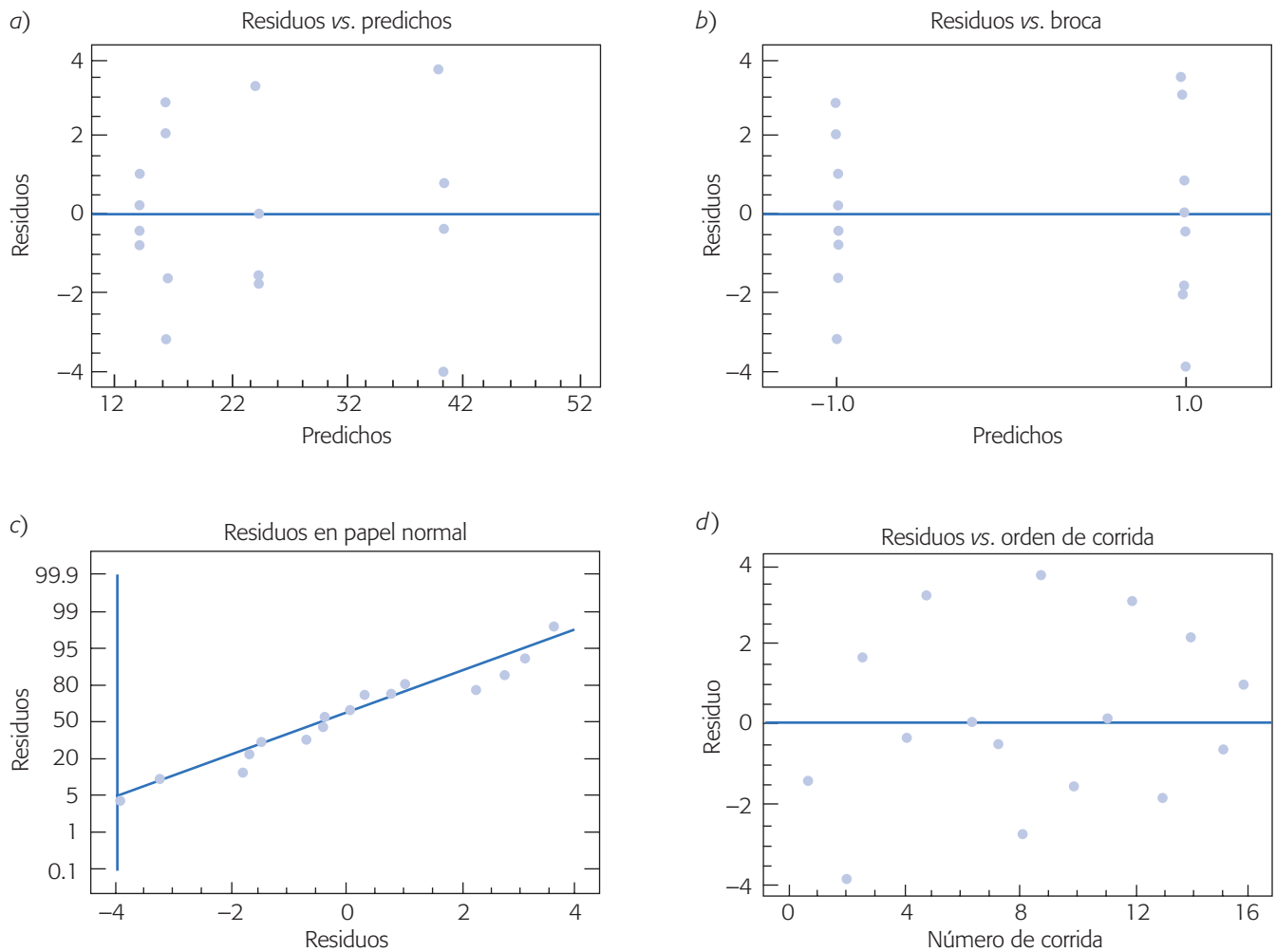


Figura 6.9 Gráfica de residuos para experimento de ranuradora.

Diseño factorial 2^3

Diseño factorial 2^3
Con este modelo se estudian tres factores en dos niveles cada uno.

Con el *diseño factorial* 2^3 se estudian *tres* factores en *dos* niveles cada uno. Consta de $2^3 = 2 \times 2 \times 2 = 8$ tratamientos diferentes, los cuales pueden identificarse con las mismas notaciones introducidas con el diseño 2^2 (véase tabla 6.1). Los tratamientos del diseño 2^3 y su representación geométrica se muestran en la figura 6.10. La región experimental ahora es un cubo regular centrado en el origen (0, 0, 0), cuyos vértices son los ocho tratamientos. La matriz de diseño se construye fácilmente alternando el signo menos y el signo más en la primera columna, dos menos y dos más en la segunda columna, y cuatro menos y cuatro más en la tercera; el diseño resulta acomodado en el *orden estándar* o de *Yates*.

Con este diseño se pueden estudiar los $2^3 - 1 = 7$ efectos: tres efectos principales A, B, C; tres interacciones dobles AB, AC, BC y una interacción triple ABC. Por lo general, el interés se enfoca en estudiar los efectos principales y las interacciones

A	B	C
−1	−1	−1
1	−1	−1
−1	1	−1
1	1	−1
−1	−1	1
1	−1	1
−1	1	1
1	1	1

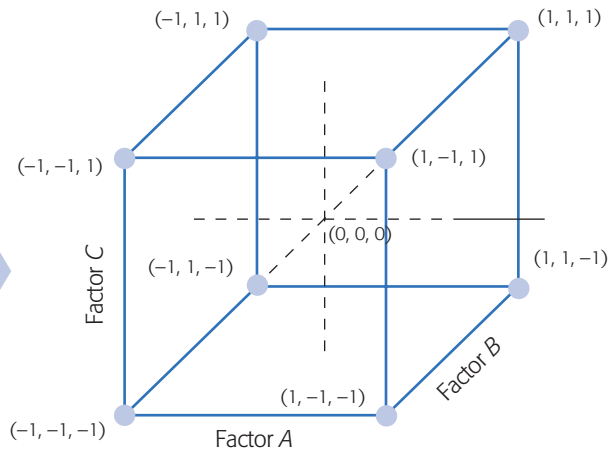


Figura 6.10 Diseño factorial 2^3 y su representación geométrica.

dobles. Sin embargo, aunque de antemano se puede considerar la interacción triple *ABC* en el diseño 2^3 como un *efecto* ignorable, es recomendable asegurarse de que su valor se mantiene pequeño, además de que al incluirla en el análisis, puede ayudar a mejorar la perspectiva de algunas gráficas, como se apreciará más adelante.

Análisis del diseño factorial 2^3

Sean *A*, *B* y *C* los factores que se quieren estudiar y sean (*I*), *a*, *b*, *ab*, *c*, *ac*, *bc* y *abc*, los totales observados en cada uno de los ocho tratamientos escritos en su orden estándar. Los *efectos* en este diseño se pueden calcular a partir de la tabla de signos (tabla 6.5).

Al igual que en el diseño 2^2, las columnas de los *efectos* principales *A*, *B* y *C* son las mismas que en la matriz de diseño, y las columnas de los *efectos* de interacción se obtienen multiplicando las columnas correspondientes.[1]

Al multiplicar las columnas de signos de la tabla 6.5 por la columna de totales representados por la notación de Yates, se obtienen los contrastes para los siete *efectos*, dados por:

$$\text{Contraste } A = [a + ab + ac + abc - (1) - b - c - bc]$$
$$\text{Contraste } B = [b + ab + bc + abc - (1) - a - c - ac]$$
$$\text{Contraste } C = [c + ac + bc + abc - (1) - a - b - ab]$$
$$\text{Contraste } AB = [ab - b - a + abc + (1) - bc - ac + c]$$
$$\text{Contraste } AC = [(1) - a + b - ab - c + ac - bc + abc]$$
$$\text{Contraste } BC = [(1) + a - b - ab - c - ac + bc + abc]$$
$$\text{Contraste } ABC = [abc - bc - ac + c - ab + b + a - (1)]$$

[1] No confundir los *efectos* de los factores con los tratamientos en la notación de Yates. Esto es, los términos *a*, *efecto A* y *factor A* representan cosas diferentes: *a* es el total o suma de las mediciones hechas en el tratamiento (+, –, –), el *efecto A* es la diferencia de medias de *Y* en los niveles alto y bajo del factor *A*, y el factor *A* es uno de los que se estudian en el experimento.

Tabla 6.5 Tabla de signos del diseño factorial 2^3.

Total	A	B	C	AB	AC	BC	ABC
(1)	−	−	−	+	+	+	−
a	+	−	−	−	−	+	+
b	−	+	−	−	+	−	+
ab	+	+	−	+	−	−	−
c	−	−	+	+	−	−	+
ac	+	−	+	−	+	−	−
bc	−	+	+	−	−	+	−
abc	+	+	+	+	+	+	+

Si se hacen n réplicas de cada tratamiento, los efectos de un diseño 2^3 se estiman dividiendo los contrastes entre $4n$. Por ejemplo, el efecto principal de A se estima de la siguiente manera:

$$\text{Efecto } A = \frac{\text{Contraste } A}{n2^{k-1}}$$

También las sumas de cuadrados de los efectos se calculan a partir de sus contrastes con la fórmula,

$$SC_{\text{efecto}} = \frac{(\text{Contraste}_{\text{efecto}})^2}{n2^k}$$

La suma total de cuadrados se obtiene de la manera usual como:

$$SC_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{l=1}^{2}\sum_{m=1}^{n} y_{ijlm}^2 - \frac{Y_{....}^2}{n2^k}$$

y, por último, la suma de cuadrados del error se calcula por sustracción. Con esta información se obtiene ANOVA para el diseño 2^3, dada en la tabla 6.6. Aquellos efectos cuyos valores-p son menores a $\alpha = 0.05$ se consideran activos y son los efectos a interpretar para conocer mejor cómo está operando el proceso y para determinar el mejor tratamiento. También recordemos que mientras menor sea el valor-p para un efecto, significa que éste tiene mayor influencia sobre la variable de respuesta.

Nótese que se requieren al menos dos repeticiones ($n \geq 2$) para calcular el cuadrado medio del error, puesto que la SC_E tiene 0 grados de libertad cuando $n = 1$. Entonces se recomienda correr este diseño con al menos dos réplicas para contar con suficientes grados de libertad para el error.

Experimento 2^3: ejemplo integrador

En una empresa que fabrica dispositivos electrónicos se identificó mediante un análisis de Pareto (Gutiérrez, 2005) que las fracturas de las obleas de silicio por choques térmicos era la principal causa de obleas rotas en las etapas de procesamiento conocidas como “grabado mesa” y “piraña”. Un grupo de esas áreas identificó a tres factores principales (temperaturas) como las probables causas del problema. Por

Tabla 6.6 ANOVA para el diseño 2^3.

FV	SC	GL	CM	F_0	Valor-p
A	SC_A	1	CM_A	CM_A/CM_E	$P(F > F_0)$
B	SC_B	1	CM_B	CM_B/CM_E	$P(F > F_0)$
C	SC_C	1	CM_C	CM_C/CM_E	$P(F > F_0)$
AB	SC_{AB}	1	CM_{AB}	CM_{AB}/CM_E	$P(F > F_0)$
AC	SC_{AC}	1	CM_{AC}	CM_{AC}/CM_E	$P(F > F_0)$
BC	SC_{BC}	1	CM_{BC}	CM_{BC}/CM_E	$P(F > F_0)$
ABC	SC_{ABC}	1	CM_{ABC}	CM_{ABC}/CM_E	$P(F > F_0)$
Error	SC_E	$2^3(n-1)$	CM_E		
Total	SC_T	$n2^3 - 1$			

ello, se utilizó un experimento factorial 2^3 con el objetivo de localizar una combinación de temperaturas en la cual se rompan un mínimo de obleas por efecto térmico.

Los tres factores controlados y sus niveles en unidades originales, son las temperaturas:

T_1 : Temperatura de grabado (–3°C, –1°C)
T_2 : Temperatura de piraña (60°C, 98°C)
T_3 : Temperatura de agua (20°C, 70°C)

La combinación (–3°C, 98°C, 20°C) fue el tratamiento usual o en operación antes del experimento. Así, uno de los dos niveles en cada factor es la temperatura usual y el otro es una temperatura que se supone reduce el efecto térmico sobre la oblea.

Tamaño de prueba. La respuesta medida a cada oblea procesada en el experimento es binaria con valor 1 si la oblea se rompe, y 0 si no se rompe. En este tipo de variables de respuesta, un asunto crítico es decidir el tamaño de prueba, es decir, el número de obleas a procesar en cada corrida. Por lo tanto, se recomienda establecerlo a partir de la estimación inicial de la magnitud del problema, de forma que en todas las corridas experimentales haya una alta probabilidad de reproducir el problema. Por ejemplo, en este caso se sabe que el número de obleas que se rompen en el tratamiento usual son 30 por cada 1 000, lo cual equivale a una proporción de $p_0 = 0.03$. La estimación del número de obleas a correr en cada prueba se calcula con la fórmula:

Tamaño de prueba o corrida
Número de piezas o unidades experimentales que se procesan en cada réplica de un tratamiento. Es importante definirla en procesos rápidos y/o con respuesta discreta.

$$m = (2.5)^2\left(\frac{1-p_0}{p_0}\right) \tag{6.13}$$

donde p_0 es la proporción utilizada como base, considerando que interesa detectar con una potencia de 90% un efecto de tamaño $0.9p_0$. En este experimento $p_0 = 0.03$ y sustituyéndolo en la relación anterior se obtiene que $m = 203$ obleas era suficiente, pero se decide utilizar $m = 250$ para detectar efectos un poco más pequeños con buena potencia. Bisgaard y Fuller (1995) proporcionan tablas para estimar el tamaño de

m, pero la regla dada en la ecuación (6.13) se ajusta bien a los tamaños de *m* que dichas tablas proporcionan. Se decide correr dos réplicas. En la tabla 6.7 se da la proporción de obleas rotas por cada 250 procesadas.

El análisis detallado del experimento se reporta en De la Vara (1994) y ahí se muestra que los datos de la tabla 6.7 se pueden analizar de diferentes maneras y se obtienen los mismos resultados. Entre esas diferentes formas está la de analizar en forma directa la proporción de obleas rotas como si fuera una variable continua, que es la solución que se describe a continuación.

Pareto estandarizado. Al aplicar las fórmulas para estimar efectos con base en los contrastes que se describieron en la sección anterior, se estima cada uno de los efectos y se obtiene la siguiente tabla:

Efecto	Estimación
A: T_Grab	–0.0195
B: T_Pira	–0.0065
C: T_Agua	–0.0085
AB	–0.0005
AC	0.0095
BC	0.0005
ABC	0.0025

Estos efectos pueden graficarse en un diagrama de Pareto para así visualizar cuáles tienen un mayor impacto sobre la variable de respuesta. Un diagrama similar al referido se muestra en la figura 6.11. Sin embargo, en ésta se muestra el

Tabla 6.7 Obleas rotas por lote y proporción por tratamiento.

T_1	T_2	T_3	*P*
–1	–1	–1	.04
1	–1	–1	.012
–1	1	–1	.036
1	1	–1	.00
–1	–1	1	.02
1	–1	1	.00
–1	1	1	.016
1	1	1	.004
–1	–1	–1	.032
1	–1	–1	.008
–1	1	–1	.028
1	1	–1	.00
–1	–1	1	.02
1	–1	1	.016
–1	1	1	.008
1	1	1	.004

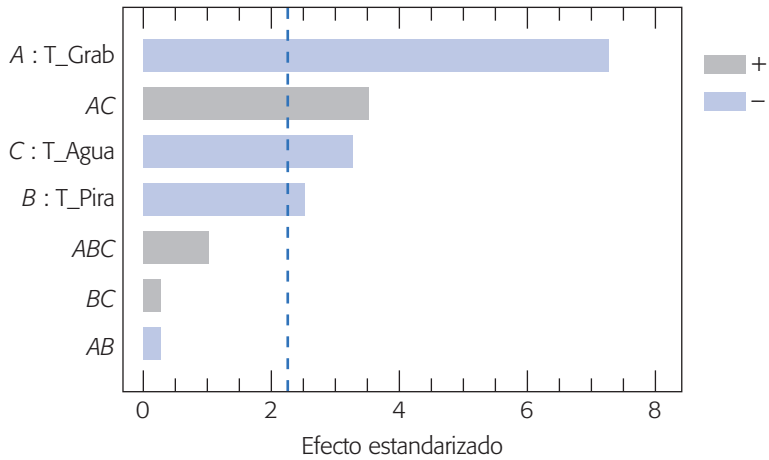


Figura 6.11 Pareto de efectos estimados para obleas.

diagrama de Pareto estandarizado, en el cual se representan los efectos divididos entre su error estándar. En general, para un diseño 2^k con n réplicas, sea CM_{error} el cuadrado medio del error que se obtiene a partir de la tabla de análisis de varianza, entonces el error estándar para un efecto puede ser estimado por:

$$\text{Estimación del error estándar de un efecto} = \hat{\sigma}_{\text{efecto}} = \sqrt{\frac{CM_{\text{error}}}{n2^{k-2}}}$$

Error estándar de un estadístico
Es la desviación estándar de un estadístico. Mide la variación muestral del mismo.

Recordemos que el *error estándar de un estadístico* es una estimación de su desviación estándar, y ésta a su vez es una estimación de la variación muestral o experimental que tiene dicho estimador. Así, en el diagrama de Pareto estandarizado se grafica la estimación de los efectos estandarizados:

$$\text{Efecto estandarizado} = \frac{\text{efecto}}{\sqrt{\frac{CM_{\text{error}}}{n2^{k-2}}}}$$

Por ejemplo, en el caso del efecto principal de A para el ejemplo de obleas rotas, tenemos que:

$$\text{Efecto estandarizado de } A = \frac{-0.0195}{\sqrt{\frac{0.000029}{2\times 2^{3-2}}}} = -7.242$$

donde el CM_{error} se obtiene de la tabla de análisis de varianza con todos los efectos incluidos (véase tabla 6.8). Los efectos estandarizados para los demás efectos se obtienen de manera similar y se representan gráficamente (en valor absoluto) en el diagrama de Pareto de la figura 6.11.

Es fácil demostrar que el efecto estandarizado sirve de estadístico de prueba para probar la hipótesis:

$$H_0 : \text{Efecto poblacional} = 0$$

contra la alternativa de que el efecto poblacional es diferente de cero. Así, se rechaza H_0 si el valor absoluto del efecto estandarizado es mayor que el valor crítico de tablas de la distribución T de Student con v grados de libertad: $t_{\alpha/2,\, v}$, donde α es el nivel de significancia prefijado para la prueba (por lo general $\alpha = 0.05$), y v son los grados de libertad asociados al error. En el ejemplo de las obleas, como el error tiene 8 grados de libertad y se trabaja con $\alpha = 0.05$, entonces de la tabla para la distribución T de Student del apéndice se obtiene que el valor crítico es $t_{025,\, 8} = 2.306$. Por lo tanto, si el valor absoluto de la estimación del efecto estandarizado es mayor que 2.306, entonces el efecto poblacional correspondiente será estadísticamente diferente de cero. Una forma sencilla de hacer esta prueba es agregar una línea en el diagrama de Pareto estandarizado a la altura del valor crítico, como se muestra en la figura 6.11. De esta manera, los efectos cuyas barras superen tal línea serán significativos. Así, para el caso de las obleas, los efectos significativos serán *A*, *AC*, *C* y *B*, en ese orden de importancia.

El mejor ANOVA. El ANOVA con todos los efectos se muestra en la tabla 6.8, ahí se aprecia que los efectos que tienen un valor-*p* menor que 0.05, son los efectos principales de *A*, *B* y *C*; y la interacción *AC* que coincide con lo visto en el diagrama de Pareto estandarizado. Además se puede notar que el valor-*p para el efecto B* está cerca de 0.05, por lo que la decisión de si tal efecto es o no significativo representa mayores riesgos de error. Con la idea de aclarar mejor cuáles fuentes de variación son significativas y obtener un modelo final en el que sólo se incluyan términos significativos, es usual construir el mejor ANOVA, en el que en una primera ronda se eliminan del análisis y se mandan al error a los efectos que claramente no son significativos. Después de esta primera ronda se revalora a los términos que estaban en una situación dudosa, como era el caso del efecto *B* en la tabla 6.8. En una segunda o tercera ronda se eliminan los términos que no resultaron significativos después de la(s) ronda(s) inicial(es).

Al hacer lo anterior en el caso del ANOVA de la tabla 6.8, es claro que los efectos *AB*, *BC* y *ABC* son no significativos, por lo que se eliminan y se mandan al error para obtener el análisis de varianza de la tabla 6.9. En ésta se observa que el efecto *B* sigue siendo significativo. De esta manera, como este ANOVA sólo tiene términos

Tabla 6.8 ANOVA completo para el ejemplo de obleas.

Efectos	SC	GL	CM	F_0	Valor-*p*
A: T_Grab	0.001521	1	0.001521	52.45	0.0001
B: T_Pira	0.000169	1	0.000169	5.83	0.0422
C: T_Agua	0.000289	1	0.000289	9.97	0.0135
AB	0.000001	1	0.000001	0.03	0.8573
AC	0.000361	1	0.000361	12.45	0.0078
BC	0.000001	1	0.000001	0.03	0.8573
ABC	0.000025	1	0.000025	0.86	0.3803
Error	0.000232	8	0.000029		
Total	0.002599	15			
$R^2 = 91.1$	$R^2_{aj} = 83.3$				

Tabla 6.9 El mejor ANOVA para el ejemplo de obleas.

Efectos	SC	GL	CM	F_0	Valor-*p*
A: T_Grab	0.001521	1	0.001521	64.60	0.0000
B: T_Pira	0.000169	1	0.000169	7.18	0.0214
C: T_Agua	0.000289	1	0.000289	12.27	0.0049
AC	0.000361	1	0.000361	15.33	0.0024
Error	0.000259	11	0.0000235		
Total	0.002599	15			
$R^2 = 90.0$	$R^2_{aj} = 86.4$				

significativos, podemos considerarlo como el mejor. También, al final de este análisis se obtiene el coeficiente de determinación que señala que un alto porcentaje de la variabilidad observada en el experimento (proporción de obleas rotas) es explicado por los efectos considerados en la tabla 6.9.

Al graficar los residuos contra *predichos* y los residuos en *papel normal* se observa el cumplimiento de los supuestos de varianza constante y normalidad.

Interpretación. De acuerdo con la figura 6.13, se aprecia una interacción importante entre la temperatura de grabado y la temperatura de agua, en particular se observa que si se trabaja con temperatura alta de grabado, prácticamente da lo mismo utilizar cualquiera de las dos temperaturas de agua. Además, en la figura 6.12 se muestran los efectos principales, por lo que de aquí y de la gráfica de interacción se concluye que hay dos condiciones de las tres temperaturas que minimizan el número de obleas rotas: (1, 1, 1) = (–1°C, 98°C, 70°C) y (1, 1, –1) = (–1°C, 98°C, 20°C). Esto también se puede apreciar en la gráfica de cubo de la figura 6.14, en la que se aprecia la respuesta predicha en cada punto del diseño. Para la predicción en cada punto del diseño (cubo) se utiliza el modelo de regresión que corresponde al mejor ANOVA (tabla 6.9), el cual, en unidades codificadas, está dado por:

$$Y = 0.01525 - 0.00975\,T_1 - 0.00325\,T_2 - 0.00425\,T_3 + 0.00475\,T_1\,T_3$$

En ese mismo cubo, comparando la respuesta predicha en la cara lateral derecha contra la cara de la izquierda del cubo, se nota que todos los tratamientos donde la temperatura de grabado (*T*-Grab) trabaja en su nivel alto, rompen menos obleas que cuando esta temperatura se encuentra en su valor bajo. La respuesta predicha en los dos puntos óptimos es cercana a 0.175% y 0.275% de obleas rotas, lo que contrasta con la respuesta en la combinación de temperaturas que se utilizaban antes del experimento, que era (T_1, T_2, T_3) = (–1, 1, –1), en donde se pronostica 3.07% de obleas rotas.

Con base en lo anterior, se decidió implementar uno de los dos mejores tratamientos y se redujo de manera significativa el número de obleas rotas por efecto térmico. Se evaluó el impacto de la mejora, y fue de 96 000 dólares anuales debido a la reducción de las obleas rotas.

Esta aplicación del diseño de experimentos muestra que para tener mejoras importantes no necesariamente se requieren diseños complicados, ni análisis estadísticos sofisticados, sino experimentos bien conducidos.

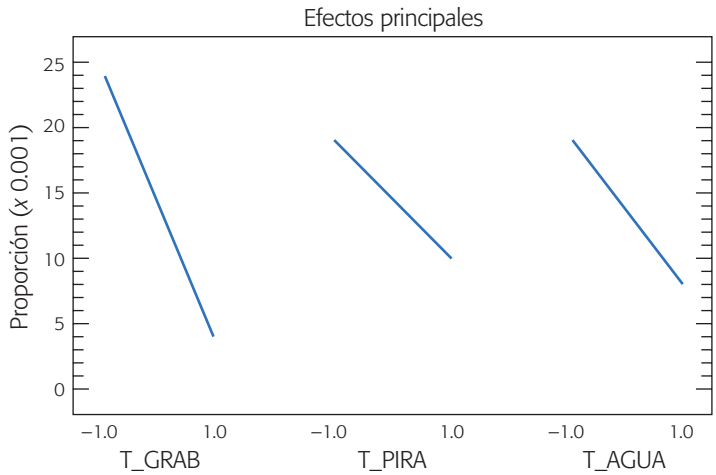


Figura 6.12 Representación de los efectos principales para obleas.

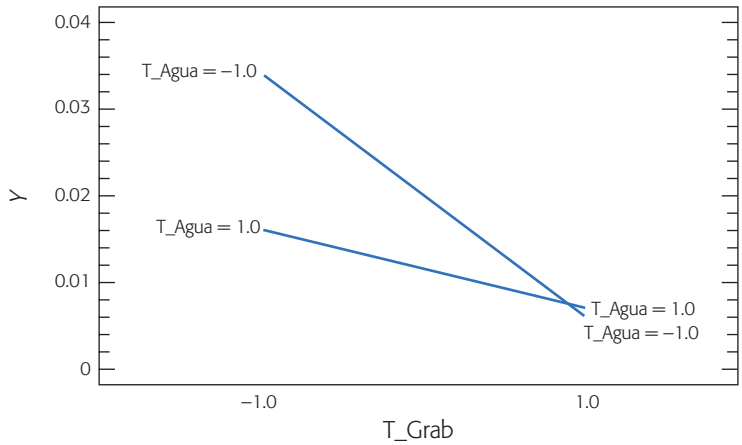


Figura 6.13 Efecto de interacción *AC* para obleas.

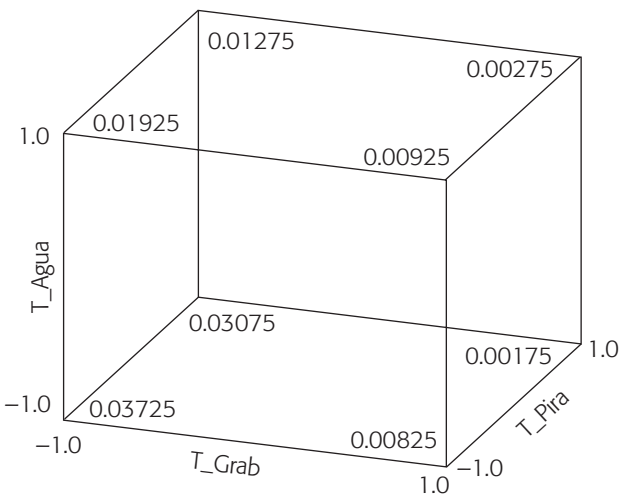


Figura 6.14 Gráfico de cubo y respuesta predicha para obleas.

Diseño factorial general 2^k

En las secciones anteriores de este capítulo hemos descrito dos casos particulares, diseños 2^2 y 2^3, del caso general que es el diseño 2^k, en el cual se consideran k factores con dos niveles cada uno, y tiene 2^k tratamientos o puntos de diseño. Las k columnas y 2^k renglones que componen la matriz para este diseño, considerando una réplica, se construyen de la siguiente manera: en la primera columna, que corresponde a los niveles del factor A, se alternan signos + y –, empezando con – hasta llegar a los 2^k renglones; en la segunda columna se alternan dos signos menos con dos signos más; en la tercera, se alternan cuatro signos menos y cuatro signos más, y así sucesivamente hasta la k-ésima columna compuesta por 2^{k-1} signos –, seguidos de 2^{k-1} signos +. En la tabla 6.10 se muestra la familia de diseños factoriales 2^k ($k \leq 5$). Nótese que el número de tratamientos siempre es potencia de dos (4, 8, 16 y 32).

Con el diseño factorial completo 2^k se pueden estudiar en total los 2^{k-1} efectos siguientes:

$$\binom{k}{1} = k, \text{ efectos principales}$$

$$\binom{k}{2} = \frac{k!}{2!(k-2)!} = \frac{k(k-1)}{2}, \text{ interacciones dobles}$$

$$\binom{k}{3} = \frac{k!}{3!(k-3)!}, \text{ interacciones triples. Y así hasta}$$

$$\binom{k}{k} = 1 \text{ interacción de los } k \text{ factores}$$

Tabla 6.10 Familia de diseños factoriales 2^k ($k \leq 5$).

Tratamiento	Notación de Yates	A	B	C	D	E
1	(1)	–	–	–	–	–
2	*a*	+	–	–	–	–
3	*b*	–	+	–	–	–
4	*ab*	+	+	–	–	–
5	*c*	–	–	+	–	–
6	*ac*	+	–	+	–	–
7	*bc*	–	+	+	–	–
8	*abc*	+	+	+	–	–
9	*d*	–	–	–	+	–
10	*ad*	+	–	–	+	–
11	*bd*	–	+	–	+	–
12	*abd*	+	+	–	+	–
13	*cd*	–	–	+	+	–
14	*acd*	+	–	+	+	–
15	*bcd*	–	+	+	+	–
16	*abcd*	+	+	+	+	–

Tratamiento	Notación de Yates	A	B	C	D	E
17	*e*	–	–	–	–	+
18	*ae*	+	–	–	–	+
19	*be*	–	+	–	–	+
20	*abe*	+	+	–	–	+
21	*ce*	–	–	+	–	+
22	*ace*	+	–	+	–	+
23	*bce*	–	+	+	–	+
24	*abce*	+	+	+	–	+
25	*de*	–	–	–	+	+
26	*ade*	+	–	–	+	+
27	*bde*	–	+	–	+	+
28	*abde*	+	+	–	+	+
29	*cde*	–	–	+	+	+
30	*acde*	+	–	+	+	+
31	*bcde*	–	+	+	+	+
32	*abcde*	+	+	+	+	+

donde la operación $\binom{k}{r} = \frac{k!}{r!(k-r)!}$ son las combinaciones de k factores tomados de r en r. Por ejemplo, el diseño factoria1 2^5 tiene 5 efectos principales, 10 interacciones dobles, 10 interacciones triples, 5 interacciones cuádruples y una interacción quíntuple, lo cual da un total $2^5 - 1 = 31$ efectos.

Estimación de contrastes, efectos y sumas de cuadrados. Cada uno de los efectos se estima a partir de su contraste, el cual a su vez se puede obtener construyendo la tabla de signos del diseño, como se explicó en las secciones anteriores. Recordemos que las *columnas de signos para los contrastes* que definen a los efectos principales están dadas directamente por la matriz de diseño, mientras que la *columna de un efecto de interacción* se obtiene multiplicando las columnas que señala dicho efecto de interacción. En la tabla 6.11 se muestra parte de la tabla de signos para un diseño 2^5.

El contraste de cada efecto se obtiene al multiplicar su columna de signos por la columna de totales expresados en la notación de Yates. Con los contrastes se procede a estimar los efectos mediante la fórmula:

$$\text{Efecto } ABC\cdots K = \frac{1}{n2^{k-1}}[\text{Contraste } ABC\cdots K] \qquad \textbf{(6.14)}$$

para el cual su suma de cuadrados con un grado de libertad está dada por:

$$SC_{AB\cdots K} = \frac{1}{n2^k}[\text{Contraste } ABC\cdots K]^2 \qquad \textbf{(6.15)}$$

donde n es el número de réplicas del experimento.

ANOVA del diseño factorial 2^k. La suma de cuadrados totales (SC_T) en el diseño factorial 2^k se calcula como:

$$SC_T = \sum_{i=1}^{n2^k} Y_i^2 - \frac{Y_\cdot^2}{n2^k}$$

y tiene $n2^k - 1$ grados de libertad, donde el subíndice i corre sobre el total de observaciones. La suma de cuadrados del error (SC_E) se obtiene por diferencia y tiene $2^k(n-1)$ grados de libertad. Con estas dos sumas de cuadrados y las de los efectos, dadas por la ecuación (6.15) se procede a escribir la tabla de ANOVA siguiendo los esquemas particulares mostrados en la tabla 6.6. Cada efecto de interés en el ANOVA es una fuente de variación para la cual se prueba la hipótesis H_0 : *efecto* = 0 *vs.* H_A : *efecto* $\neq$ 0. Así, cuando se concluye que un efecto está activo, significa que es estadísticamente diferente de cero.

Si en la tabla de ANOVA se incluye el total de efectos que se estiman con el factorial completo 2^k, será necesario realizar cuando menos dos réplicas del experimento para estimar una suma de cuadrados del error. Sin embargo, en la mayoría de los casos sólo interesa estudiar los efectos principales y las interacciones dobles. Esto hace que cuando el número de factores es mayor o igual a cuatro ($k \geq 4$) no sea estrictamente necesario realizar réplicas. En la siguiente sección se verá esto.

Cabe agregar que cuando se emplea un diseño factorial 2^k, se supone que la respuesta es aproximadamente lineal en el rango de variación de cada uno de los

Tabla 6.11 Diseños factoriales 2^k y sus efectos de interés, $k \leq 5$.

Notación de Yates	A	B	C	D	E	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
(1)	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+
a	+	−	−	−	−	−	−	−	−	+	+	+	+	+	+
b	−	+	−	−	−	−	+	+	+	−	−	−	+	+	+
ab	+	+	−	−	−	+	−	−	−	−	−	−	+	+	+
c	−	−	+	−	−	+	−	+	+	−	+	+	−	−	+
ac	+	−	+	−	−	−	+	−	−	−	+	+	−	−	+
bc	−	+	+	−	−	−	−	+	+	+	−	−	−	−	+
abc	+	+	+	−	−	+	+	−	−	+	−	−	−	−	+
d	−	−	−	+	−	+	+	−	+	+	−	+	−	+	−
ad	+	−	−	+	−	−	−	+	−	+	−	+	−	+	−
bd	−	+	−	+	−	−	+	−	+	−	+	−	−	+	−
abd	+	+	−	+	−	+	−	+	−	−	+	−	−	+	−
cd	−	−	+	+	−	+	−	−	+	−	−	+	+	−	−
acd	+	−	+	+	−	−	+	+	−	−	−	+	+	−	−
bcd	−	+	+	+	−	−	−	−	+	+	+	−	+	−	−
abcd	+	+	+	+	−	+	+	+	−	+	+	−	+	−	−
e	−	−	−	−	+	+	+	+	−	+	+	−	+	−	−
ae	+	−	−	−	+	−	−	−	+	+	+	−	+	−	−
be	−	+	−	−	+	−	+	+	−	−	−	+	+	−	−
abe	+	+	−	−	+	+	−	−	+	−	−	+	+	−	−
ce	−	−	+	−	+	+	−	+	−	−	+	−	−	+	−
ace	+	−	+	−	+	−	+	−	+	−	+	−	−	+	−
bce	−	+	+	−	+	−	−	+	−	+	−	+	−	+	−
abce	+	+	+	−	+	+	+	−	+	+	−	+	−	+	−
de	−	−	−	+	+	+	+	−	−	+	−	−	−	−	+
ade	+	−	−	+	+	−	−	+	+	+	−	−	−	−	+
bde	−	+	−	+	+	−	+	−	−	−	+	+	−	−	+
abde	+	+	−	+	+	+	−	+	+	−	+	+	−	−	+
cde	−	−	+	+	+	+	−	−	−	−	−	−	+	+	+
acde	+	−	+	+	+	−	+	+	+	−	−	−	+	+	+
bcde	−	+	+	+	+	−	−	−	−	+	+	+	+	+	+
abcde	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

factores estudiados. No es necesario suponer una linealidad perfecta, pero sí que no haya una curvatura muy grande. De esta manera, dado que cada factor se prueba en dos niveles, no es posible estudiar efectos de curvatura (efectos del tipo A^2, B^2, etc.), aunque ésta exista en el proceso; para estudiar tales efectos se necesitan al menos tres niveles en cada factor. Esto no implica que de entrada sea recomendable un diseño factorial con al menos tres niveles en cada factor, sino que en primera instancia se pueden agregar repeticiones (mínimo tres) al centro del diseño factorial 2^k, y con ellas se podrá detectar la presencia de curvatura. Los factoriales con punto central se

discuten más adelante en este capítulo y la metodología para estudiar la curvatura se presenta en el capítulo 12.

Diseño factorial 2^k no replicado

Número de réplicas en los factoriales 2^k

Al aumentar el número de factores en el diseño 2^k crece rápidamente el número de tratamientos y, por lo tanto, el número de corridas experimentales. Si se realizan dos repeticiones en cada punto se tienen que hacer 2×2^k corridas experimentales, lo cual consume muchos recursos para cinco o más factores ($k \geq 5$), ya que serían 64 o más corridas. El diseño 2^4 es quizás el factorial más grande que todavía se puede correr con dos réplicas, lo que implica hacer 32 corridas, pero incluso con una réplica (16 pruebas) de este diseño muchas veces es suficiente para estudiar los 10 efectos de interés (tabla 6.12). De acuerdo con esta tabla, cuando se trata de cuatro factores se recomienda, en primera instancia, correr una sola vez el diseño; cuando son cinco factores se recomienda correr sólo la mitad del diseño (fracción 2^{5-1}), y después de analizar esta primera mitad se decide completar una réplica del factorial 2^5 completo; de seis factores en adelante, el diseño siempre se corre fraccionado y sólo una réplica de la fracción elegida. En el capítulo 8 se estudian con detalle los diseños factoriales fraccionados.

Nótese que ninguno de los diseños listados en la tabla tiene más de 32 corridas. Se puede afirmar que la mayoría de los experimentos factoriales 2^k o las fracciones de ellos que se utilizan en la práctica, requieren a lo más 32 corridas experimentales, y con ellas se puede estudiar hasta una cantidad grande de factores ($k > 8$).

Más aún, un máximo de 16 pruebas son suficientes para la mayoría de los problemas en una primera etapa de experimentación.

Una sola réplica o corrida del factorial 2^k completo es una estrategia adecuada cuando se tienen cuatro o más factores, considerando que a partir de $k = 4$ se comienza a tener mucha información con el diseño factorial completo. Por ejemplo, en el caso de $k = 5$, los efectos se estimarían como la diferencia de medias de 16 datos cada una. Pero además se puedan estimar interacciones de alto orden, que por lo general no son significativas. Tales interacciones pueden utilizarse para estimar un error que permita construir un ANOVA aproximado. En el diseño factorial 2^5 una repetición es suficiente para estimar sus $2^5 - 1 = 31$ efectos totales. De estos efectos se pueden ignorar de antemano las 16 interacciones de tres o más factores, y utilizarlos para

Tabla 6.12 Réplicas o corridas en la familia de diseños 2^k.

Diseño	Réplicas recomendadas	Número de corridas
2^2	3 o 4	12, 16
2^3	2	16
2^4	1 o 2	16, 32
2^5	fracción 2^{5-1} o 1	16, 32
2^6	fracción 2^{6-2} o fracción 2^{6-1}	16, 32
2^7	fracción 2^{7-3} o fracción 2^{7-2}	16, 32

construir un error aproximadamente aleatorio, cuya suma de cuadrados del error (SC_E) sería la suma de las sumas de cuadrados de los efectos mandados al error, los grados de libertad de (SC_E) son tantos como los efectos que se aglomeran para conformar dicho error.

La construcción de la suma de cuadrados del error (SC_E) para el ANOVA de un factorial 2^k con una sola réplica se realiza a través de los dos pasos siguientes:

1. Se puede suponer de antemano que las interacciones de tres o más factores no son significativas y enviadas directamente al error. Sin embargo, es recomendable que antes de enviar al error las interacciones triples se verifiquen, mediante técnicas gráficas, que efectivamente son efectos despreciables. Estas técnicas gráficas se describen en la siguiente subsección.
2. Se utilizan técnicas gráficas, tabulares y numéricas para decidir cuáles de los efectos principales, interacciones dobles y triples se pueden enviar al error. Con los efectos excluidos se obtiene una suma de cuadrados del error, y con ella se construye la tabla de análisis de varianza. Este ANOVA es sólo una aproximación, ya que siempre existe el riesgo de que la magnitud del error así construido no sea la correcta, dado que no se basa en repeticiones auténticas. Una manera de saber si el cuadrado medio del error (CM_E) resultante es apropiado consiste en compararlo con la varianza σ^2 típica que haya observado la respuesta en su comportamiento previo al experimento. Recordemos que el cuadrado medio del error es un estimador de la varianza.

Se deben buscar, eliminar o mandar al error al menos ocho efectos pequeños para que tenga mayores posibilidades de estar bien estimados. Puede haber efectos con los que no es nada claro si se deben o no mandar el error. Con estos efectos la decisión se debe basar en todos los argumentos aplicables que se describen enseguida.

¿Cómo decidir cuáles efectos mandar al error?

Cuando se corre sólo una réplica del experimento, el cuadrado medio del error (CM_E), necesario para probar la significancia de cada efecto, debe construirse a partir de efectos pequeños o despreciables. Existen varias técnicas que ayudan a detectar con bastante seguridad y sin la necesidad de un ANOVA, los efectos que pueden utilizarse para conformar el error.

El arte en el uso de tales técnicas es lo que permite construir un cuadrado medio del error lo más apegado posible a la realidad. Si mandamos al error un efecto que no se debe al azar, que es un efecto real, éste puede inflar el CM_E reduciendo la potencia del ANOVA para detectar efectos significativos (sería como echar una cortina de humo que no dejara ver qué sucede con los efectos); por otro lado, si el error resulta exageradamente pequeño, se estarían detectando como significativos efectos que no lo son, lo cual puede llevar a decisiones incorrectas. Si el cuadrado medio del error resulta muy diferente de la σ^2 histórica de la misma respuesta, es un síntoma de que posiblemente no está bien estimado.

En cualquier experimento pueden existir tres tipos de efectos: los que claramente afectan, los que claramente no afectan y los efectos intermedios. Cuando exis-

ten sólo efectos de los dos primeros es fácil decidir con cuáles conformar el error. El problema es que cuando existen efectos intermedios, no está nada claro si afectan o no a la respuesta.

Las técnicas para decidir qué efectos mandar al error no funcionan bien cuando los efectos que tiene el diseño son pocos. Como es el caso de los diseños 2^2 y 2^3. Pero, con frecuencia, estos diseños se corren con réplicas suficientes para construir el análisis de varianza directamente (tabla 6.12). Las técnicas que se describen a continuación tienen mayor utilidad en diseños con cuatro factores en adelante, en los que existen 10 o más efectos a investigar. Estas técnicas se ejemplifican en la siguiente sección.

Gráfico de efectos en papel normal (Gráfica de Daniel)

Gráfico de efectos en papel normal
Gráfico que permite visualizar cuáles efectos pueden ser significativos: entre más se aleje un punto de la línea, más importante será el correspondiente efecto.

Al usar los efectos como sumas de variables aleatorias (diferencia de medias), Daniel (1959) se dio cuenta de que los efectos no significativos deben seguir una distribución normal con media igual a cero y varianza constante. Esto implica que si los efectos se grafican en *papel probabilístico normal*, los que no son significativos tenderán a formar una línea recta, mientras que los efectos activos aparecerán alejados de la línea de normalidad.

Papel probabilístico medio normal
Gráfica de efectos basada sólo en la parte positiva de la distribución normal. Cuando se tienen efectos positivos y negativos da una mejor perspectiva de los que pueden ser significativos.

Cuando se tienen efectos positivos y negativos es mejor utilizar el *papel probabilístico medio normal* (*half normal*), para tener una mejor perspectiva de cuáles efectos se alinean y cuáles no. Como su nombre lo indica, el papel medio normal sólo utiliza la parte positiva de la distribución normal estándar, aprovechando su simetría y el hecho de que dos efectos de signo contrario y de la misma magnitud son igualmente importantes.

Como se vio anteriormente, el papel probabilístico normal también sirve para verificar el cumplimiento del supuesto de normalidad de los residuos. La gráfica de efectos en papel normal tiene un objetivo muy diferente a esta gráfica de residuos.

Diagrama de Pareto de efectos

Diagrama de Pareto de efectos
Gráfico de barras que representa los efectos ordenados en forma descendente de acuerdo con su magnitud absoluta.

El *diagrama de Pareto* para los efectos sin estandarizar representa una manera práctica de ver cuáles efectos son los más grandes en cuanto a su magnitud. El Pareto representa de manera descriptiva la realidad observada de los efectos, pero sin considerar supuestos distribucionales. En la gráfica de efectos en papel de probabilidad normal (gráfico de Daniel) es más difícil apreciar la importancia relativa de los efectos, pero es mejor que el Pareto para señalar cuáles efectos son activos. Por ello, lo mejor es utilizar ambas gráficas para decidir cuáles efectos mandar al error.

Con el Pareto y el gráfico de Daniel muchas veces se logran detectar claramente los efectos significativos, y una vez que se construya el error, el ANOVA sólo confirmará lo que ya se ha encontrado con estos gráficos. Se dice que el diagrama de Pareto trabaja limpiamente cuando quedan bien delimitados los diferentes grupos de efectos, de los más a los menos importantes (véase figura 6.15*a*). En esta figura, cada concavidad de la línea sobrepuesta a las barras indica las oleadas o rachas que ocurren, y en este caso básicamente habría dos posibilidades para construir el error y hacer el análisis de varianza: excluir el primer grupo de menor importancia o también se excluye el segundo grupo de menor importancia. Por otra parte, si las barras

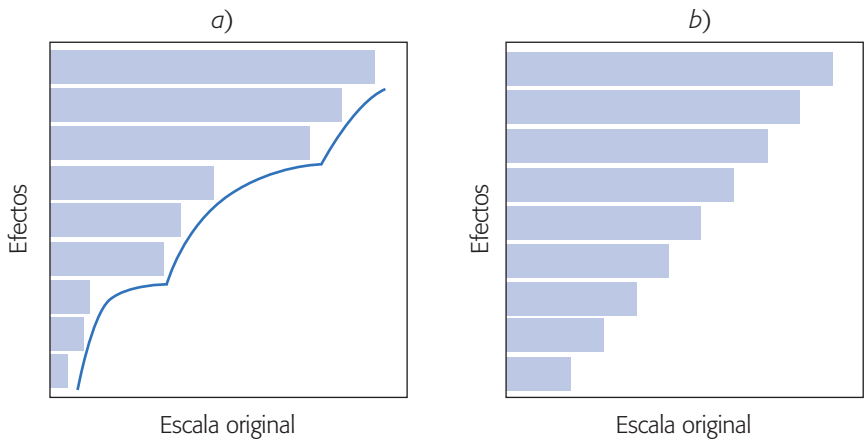


Figura 6.15 Diagramas de Pareto que funcionan diferente: *a*) limpiamente, y *b*) con el Pareto no es obvio dónde hacer el corte.

del diagrama quedan como escalones de igual tamaño (véase figura 6.15*b*), el principio de Pareto no está trabajando limpiamente, y en esta situación es necesario usar otros criterios que ayuden a dilucidar dónde hacer el corte de exclusión.

Otros criterios útiles

En ocasiones, ni el gráfico de Daniel ni el diagrama de Pareto aclaran bien la situación de algunos efectos de magnitud intermedia. En estos casos, para decidir cuáles de estos efectos se mandarán al error, se recomienda fijarse en todos los criterios siguientes y no sólo en uno de ellos:

1. La magnitud del efecto. Si se conoce la desviación estándar σ del proceso, la magnitud del efecto puede indicar si éste se manda al error. De manera específica, en el factorial 2^k con una réplica se compara el efecto observado contra dos veces el error estándar del efecto $(\sigma/\sqrt{2^{k-2}})$ y si el primero es más grande es porque puede ser un efecto real.
2. Si primero se excluyen los efectos que son claramente no significativos de acuerdo al gráfico de Daniel y al Pareto, se puede lograr un ANOVA preliminar cuya significancia da información útil para excluir o no los efectos restantes. Específicamente, los efectos cuyas significancias en el ANOVA preliminar están alrededor de 0.2 o menores, no necesariamente se excluyen del análisis. Esta decisión es más confiable cuando dicho ANOVA preliminar ya alcanzó al menos 8 grados de libertad para el error.
3. Los grados de libertad del error deben ser al menos 8 para tener un ANOVA más confiable.
4. El R^2_{aj} del modelo en el ANOVA preliminar. Cuando se van eliminando efectos que no son significativos, el estadístico R^2_{aj} crece. En el momento en que se elimina un efecto y este estadístico decrece 3% o más, significa que posiblemente ese efecto no debe excluirse.

Colapsación o proyección del diseño

Cuando en el mejor ANOVA que se pudo determinar se detecta que un factor particular no es significativo, ya que su efecto principal y todas las interacciones en las que interviene no son importantes, entonces en lugar de mandar al error este factor y sus interacciones, otra posibilidad es *colapsar* o *proyectar el diseño*, lo cual consiste en eliminar completamente del análisis a tal factor, con lo que el diseño factorial 2^k original se convierte en un diseño completo con un factor menos (2^{k-1}) y con dos repeticiones en cada punto. Al haber "repeticiones" en el diseño 2^{k-1} resultante de la colapsación, entonces se puede estimar el CM_E y construir la tabla de análisis de varianza de la manera usual. En general, si se pueden omitir h factores, los datos se convierten en un diseño factorial 2^{k-h} con 2^h repeticiones en cada punto. Por ejemplo, si se eliminan dos factores, el diseño 2^k con una réplica, entonces el diseño resultante es un diseño factorial completo con $k - 2$ factores y cuatro réplicas.

Proyectar el diseño
Consiste en eliminar por completo un factor del análisis, con lo que el diseño 2^k original se convierte en un diseño con un factor menos (2^{k-1}) y con el doble de repeticiones en cada tratamiento.

El efecto de colapsar un diseño factorial 2^3 se representa en la figura 6.16. Es como construir un cubo de cartón que representa el diseño 2^3, colocarlo en el piso cuidando que el factor que no afecta quede en el sentido vertical, y hacer fuerza sobre el cubo hasta que éste se colapse y se convierta en un plano. El resultado es un diseño factorial 2^2 con el doble de réplicas que tenía el diseño original. Si se tenía una réplica, al colapsar un factor se tienen dos réplicas. Debemos aclarar que la acción de colapsar facilita la reproducción posterior del análisis del experimento, al eliminar del análisis los factores que se sabe que no tienen influencia.

Experimento 2^5 no replicado: ejemplo integrador

En una planta donde se fabrican semiconductores se quiere mejorar el rendimiento del proceso vía diseño de experimentos. De acuerdo con la experiencia del grupo de mejora, los factores que podrían tener mayor influencia sobre la variable de respuesta (rendimiento), así como los niveles de prueba utilizados son los siguientes:

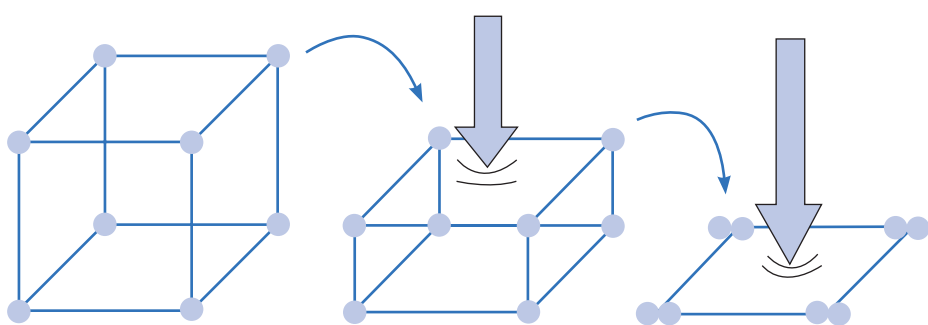


Figura 6.16 Acción de colapsar un factorial 2^3.

A = Nivel de la abertura (pequeña, grande).
B = Tiempo de exposición (20% abajo, 20% arriba).
C = Tiempo de revelado (30 seg, 45 seg).
D = Dimensión de la máscara (pequeña, grande).
E = Tiempo de grabado (14.5 min, 15.5 min).

Se decide correr un experimento 2^5 con una sola corrida o réplica para estudiar estos cinco factores. En la tabla 6.13 se muestra la matriz de diseño con los tratamientos en orden aleatorio. Esta matriz debe guardarse en un archivo para no perder ese orden y capturar los datos una vez hechas las pruebas. En la tabla se muestran los primeros tres valores observados del rendimiento. A la hora de correr el experimento es mejor llevar escrita la matriz de diseño en las unidades originales, para facilitar la operación del proceso bajo cada condición. En este caso hemos usado unidades codificadas por razones de espacio.

Se hacen las 32 corridas a nivel proceso, indicadas en la tabla anterior. Los resultados escritos en el orden estándar con la notación de Yates se muestran en la tabla 6.14.

Análisis del experimento

Lo primero es estimar los efectos potencialmente importantes: cinco efectos principales, 10 interacciones dobles y 10 interacciones triples. Todos los efectos de las interacciones triples son pequeños; por ello, en la tabla 6.15 sólo hemos reportado los efectos principales y las interacciones dobles. Éstos se analizan con el Pareto para efectos y la gráfica de Daniel (figuras 6.17 y 6.18). Llaman la atención los cuatro

Tabla 6.13 Matriz de diseño en orden aleatorio.

Corrida	A	B	C	D	E	Y
1	1.0	1.0	–1.0	–1.0	–1.0	55
2	–1.0	1.0	1.0	1.0	–1.0	44
3	1.0	1.0	1.0	1.0	–1.0	61
4	–1.0	–1.0	–1.0	–1.0	–1.0	–
5	–1.0	1.0	–1.0	1.0	–1.0	–
6	1.0	–1.0	1.0	–1.0	–1.0	–
7	–1.0	1.0	1.0	–1.0	1.0	–
8	–1.0	–1.0	–1.0	1.0	–1.0	–
9	1.0	1.0	–1.0	–1.0	1.0	–
10	–1.0	–1.0	1.0	–1.0	1.0	–
11	–1.0	1.0	1.0	–1.0	–1.0	–
12	–1.0	1.0	–1.0	–1.0	1.0	–
13	–1.0	–1.0	1.0	1.0	–1.0	–
14	1.0	1.0	1.0	1.0	1.0	–
15	1.0	–1.0	1.0	1.0	1.0	–
16	1.0	1.0	–1.0	1.0	1.0	–

Corrida	A	B	C	D	E	Y
17	–1.0	–1.0	–1.0	1.0	1.0	–
18	–1.0	–1.0	–1.0	–1.0	1.0	–
19	1.0	–1.0	–1.0	–1.0	1.0	–
20	–1.0	–1.0	1.0	–1.0	–1.0	–
21	1.0	–1.0	1.0	–1.0	1.0	–
22	1.0	1.0	–1.0	1.0	–1.0	–
23	1.0	–1.0	–1.0	1.0	–1.0	–
24	–1.0	–1.0	1.0	1.0	1.0	–
25	1.0	1.0	1.0	–1.0	–1.0	–
26	1.0	–1.0	–1.0	1.0	1.0	–
27	–1.0	1.0	–1.0	1.0	1.0	–
28	–1.0	1.0	–1.0	–1.0	–1.0	–
29	–1.0	1.0	1.0	1.0	1.0	–
30	1.0	1.0	1.0	–1.0	1.0	–
31	1.0	–1.0	1.0	1.0	–1.0	–
32	1.0	–1.0	–1.0	–1.0	–1.0	–

Tabla 6.14 Datos acomodados en el orden estándar.

(1) = 7	*d* = 8	*e* = 18	*de* = 6
a = 9	*ad* = 10	*ae* = 12	*ade* = 10
b = 34	*bd* = 32	*be* = 35	*bde* = 30
ab = 55	*abd* = 50	*abe* = 52	*abde* = 53
c = 16	*cd* = 18	*ce* = 15	*cde* = 15
ac = 20	*acd* = 21	*ace* = 22	*acde* = 20
bc = 40	*bcd* = 44	*bce* = 45	*bcde* = 41
abc = 60	*abcd* = 61	*abce* = 65	*abcde* = 63

efectos cuya magnitud es bastante grande en comparación con la de los demás efectos (*A*, *B*, *C*, *AB*).

En el análisis de varianza preliminar de la tabla 6.16 también se han mandado al error las interacciones de tres factores en adelante, esto permite generar 16 grados de libertad del error, lo cual es el doble de lo mínimo recomendado para los grados de libertad para el error. Recordemos que si se incluyen los 31 efectos en el análisis, el ANOVA quedaría incompleto porque no habría grados de libertad para el error. El ANOVA de la tabla 6.16 parece adecuado, en el sentido de que separa bastante bien a los cuatro efectos más importantes que ya se han mencionado. Podemos ver que, aun sin ser el mejor ANOVA, puesto que incluye bastantes términos que no influyen, son sólo cuatro los efectos que tienen un valor-*p* bastante más chico que el valor de $\alpha = 0.05$. Así, este ANOVA preliminar hace un buen trabajo al determinar cuáles efectos son significativos. Quizá la duda pudiera ser la interacción *DE*, con valor-*p* = 0.049. De este ANOVA se puede estimar provisionalmente el error estándar de los efectos, que como ya hemos visto, está dado por:

$$\hat{\sigma}_{\text{efecto}} = \sqrt{\frac{CM_{\text{error}}}{n2^{k-2}}} = \sqrt{\frac{2.48}{1 \times 2^{5-2}}} = 0.557$$

y recordemos que si un efecto es más grande que el doble del error estándar, es síntoma de que puede ser importante. De acuerdo con la tabla 6.15, en este caso sólo caen los efectos *A*, *B*, *C*, *AB*; y seguiría en duda la interacción *DE*.

Mejor ANOVA. Con la idea de despejar dudas y llegar al mejor ANOVA, lo que sigue es eliminar los efectos menos importantes. Los gráficos de Pareto y de Daniel muestran claramente que sólo hay cuatro efectos significativos. En la figura 6.17 se observa que el gráfico de Pareto hace un trabajo limpio al detectar a los cuatro efectos

Mejor ANOVA
ANOVA del modelo más simple que explica mejor el comportamiento de la variable de respuesta. Se obtiene eliminando los términos que no contribuyen.

Tabla 6.15 Efectos estimados.

A: Abertu = 11.8125	*AE* = 0.9375
B: T-expo = 33.9375	*BC* = 0.0625
C: T-revel = 9.6875	*BD* = –0.6875
D: máscara = –0.8125	*BE* = 0.5625
E: T-grab = 0.4375	*CD* = 0.8125
AB = 7.9375	*CE* = 0.3125
AC = 0.4375	*DE* = –1.1875
AD = –0.0625	

Tabla 6.16 ANOVA preliminar para los semiconductores.

FV	SC	GL	CM	F_0	Valor-*p*
A: Abertu	1 116.28	1	1 116.28	449.32	0.0000
B: T-expo	9 214.03	1	9 214.03	3 708.79	0.0000
C: T-revel	750.78	1	750.78	302.20	0.0000
D: máscara	5.28	1	5.28	2.13	0.1642
E: T-grab	1.53	1	1.53	0.62	0.4439
AB	504.03	1	504.3	202.88	0.0000
AC	1.53	1	1.53	0.62	0.4439
AD	0.03	1	0.03	0.01	0.9121
AE	7.03	1	7.03	2.83	0.1119
BC	0.03	1	0.03	0.01	0.9121
BD	3.78	1	3.78	1.52	0.2351
BE	2.53	1	2.53	1.02	0.3278
CD	5.28	1	5.28	2.13	0.1642
CE	0.78	1	0.78	0.31	0.5827
DE	11.28	1	11.28	5.54	0.0490
Total error	39.75	16	2.48		
Total	11 664.0	31			

que están activos, así como a los efectos que deben eliminarse para obtener el mejor y el definitivo análisis de varianza.

Las mismas cuatro barras más grandes en el diagrama de Pareto corresponden a los cuatro puntos (efectos) que se alejan de la línea en el papel normal (véase figura 6.18) y que es señal de que son efectos reales en el proceso. Los efectos alineados se deben al azar, por ello se pueden mandar al error. Note que en este ejemplo todo lo que sucede con los efectos es claro desde el principio, cosa que no siempre ocurre.

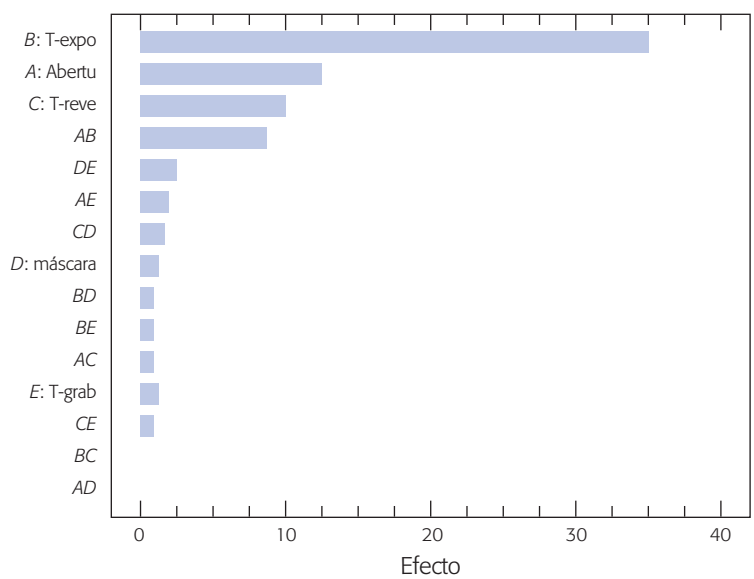


Figura 6.17 Pareto de efectos para ejemplo de semiconductores.

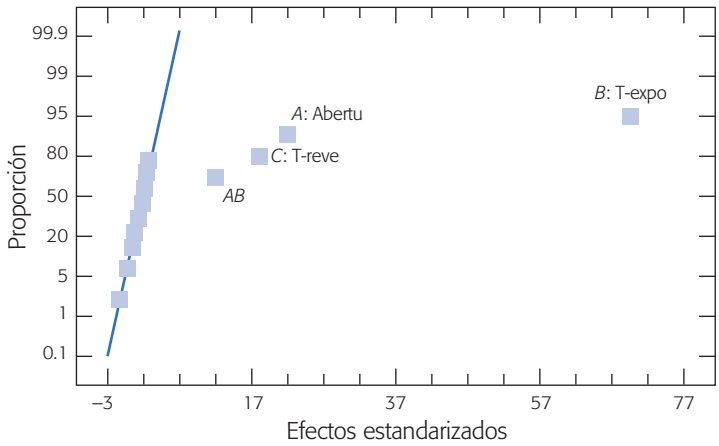


Figura 6.18 Efectos en papel normal (Gráfica de Daniel).

Al eliminar los efectos indicados se obtiene el mejor análisis de varianza dado en la tabla 6.17. Este análisis es el mejor porque, además de que detecta de manera contundente a los efectos significativos, es el modelo más simple posible para explicar el comportamiento del rendimiento durante las pruebas. En el ANOVA preliminar (véase tabla 6.16) también aparecía la interacción *DE* como un efecto significativo (valor-$p < 0.05$), y si la incluyéramos en este mejor análisis seguiría apareciendo como significativa. Sin embargo, no se debe incluir porque su aportación a la explicación del rendimiento es mínima: estos cuatro efectos explican, según el estadístico R^2_{aj}, 99.22% de la variabilidad observada, y si se incluye en el ANOVA el efecto *DE*, el estadístico sube a 99.33%, es decir, es un aumento muy pequeño de 0.10%. Esto muestra que, aunque tal efecto resulta significativo, es en realidad espurio. En otras palabras, es más lo que estorba el incluido que lo que ayuda, y puede haber más ganancia si se deciden los niveles de los factores *D* y *E* al utilizar un criterio económico, que con base en su interacción.

Debemos decir que el mejor ANOVA no es único puesto que puede intervenir bastante el criterio del experimentador y, a partir de los mismos datos, dos personas podrían llegar a dos mejores ANOVA un poco diferentes. No obstante, en su parte medular ambos ANOVA deben coincidir.

Tabla 6.17 El mejor análisis de varianza.

FV	*SC*	*GL*	*CM*	F_0	Valor-*p*
A: Abertu	1 116.28	1	1 116 .28	382.27	0.0000
B: Texpo	9 214.03	1	9 214.03	3 155.34	0.0000
C: Teve	750.78	1	750.78	257.10	0.0000
AB	504.03	1	504.03	172.61	0.0000
Error	78.84	27	2.92		
Total	11 664.0	31			
$R^2_{ij} = 99.22\%$					

Interpretación

Se procede a interpretar los cuatro efectos que resultaron significativos en el mejor análisis de varianza de la tabla 6.17, que son: *A*, *B*, *C* y *AB*. Los tres efectos principales se muestran en la figura 6.19, y de ellos sólo se interpreta el efecto de *C* (tiempo de revelado) puesto que *A* y *B* interactúan, lo cual tiene prioridad. Recordando que la variable de respuesta es rendimiento, de la gráfica del efecto *C* se concluye que a mayor el tiempo de revelado mayor es el rendimiento; por lo tanto, el tiempo de revelado debe fijarse en C^+.

El efecto de interacción *AB* se muestra en la figura 6.20. Se observa que el efecto del factor *A* es mayor cuando el factor *B* está en más; además, el extremo de línea más alto en la escala del rendimiento corresponde claramente a la combinación (A^+, B^+), es decir, la abertura (*A*) debe estar en su tamaño grande y es mejor el mayor tiempo de exposición (*B*). Es importante reflexionar y analizar las razones físicas y de ingeniería, de por qué estos efectos influyen de tal manera sobre *Y*, con lo que además de encontrar soluciones se estaría generando conocimiento.

En conclusión, el mejor tratamiento es (A^+, B^+, C^+, $D^\$$, $E^\$$); *A*, *B* y *C* en su nivel alto y para *D* y *E* se eligen los niveles que resultan más convenientes desde el punto de vista económico, de productividad o de operabilidad. En este caso se decide utilizar la dimensión menor de la máscara (D^-) y el menor tiempo de grabado (E^-). *Otro* criterio es elegir los niveles de estos factores donde la variabilidad del rendimiento sea menor.

Predicción. Para predecir el rendimiento esperado en el mejor tratamiento (A^+, B^+, C^+, D^-, E^-) o en la combinación que se quiera, se obtiene el modelo de regresión ajustado relacionado al mejor ANOVA, que está dado por:

$$\hat{Y} = 30.53 + 5.91x_1 + 16.97x_2 + 5.84x_3 + 3.97x_1x_2 \qquad \textbf{(6.16)}$$

donde x_1 es el factor *A*, x_2 es el factor *B* y x_3 el factor *C*; $\hat{Y}$ es el rendimiento predicho en el punto (x_1, x_2, x_3) en unidades codificadas, como en la tabla 6.13. Recordemos que, al utilizar datos codificados, los coeficientes del modelo son iguales a la mitad del efecto estimado correspondiente, representado en la tabla 6.16. En la gráfica de

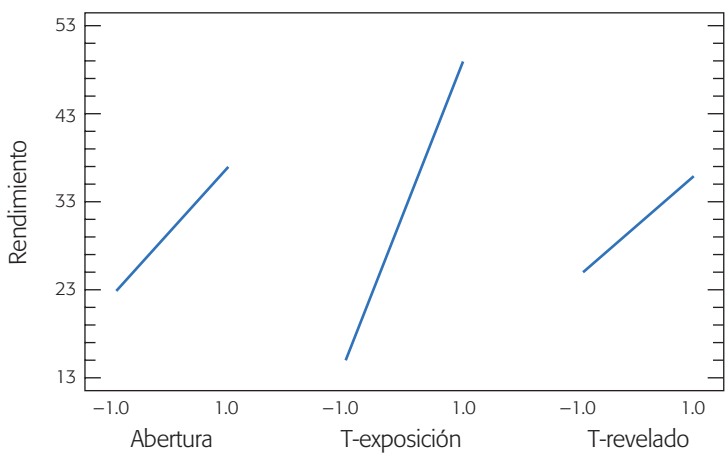


Figura 6.19 Efectos principales: *A, B* y *C*; ejemplo de semiconductores.

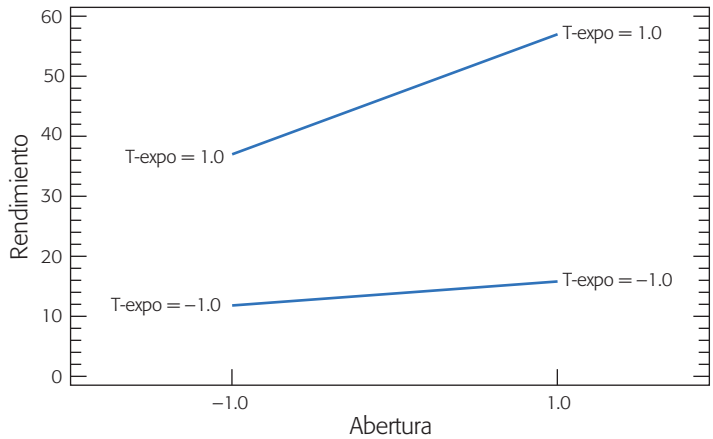


Figura 6.20 Efecto de interacción *AB*; ejemplo de semiconductores.

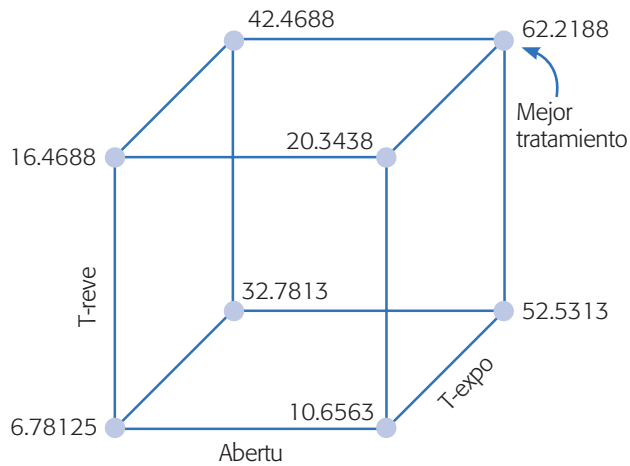


Figura 6.21 Gráfica de cubo y respuesta predicha; ejemplo de semiconductores.

cubo de la figura 6.21 se reporta el rendimiento predicho por el modelo en cada combinación de los tres factores incluidos. En particular, en el mejor tratamiento se predice un rendimiento promedio de 62.22%.

Verificación de supuestos

La *verificación de supuestos* debe hacerse para el modelo que corresponde al mejor ANOVA. De violarse alguno de los supuestos, el análisis podría dar conclusiones incorrectas. Debemos observar que en este ejemplo la situación con los efectos es tan contundente que aun una violación clara de los supuestos difícilmente cambiaría las conclusiones. En la figura 6.22 se grafican los predichos contra los residuos, y se muestra una ligera violación al supuesto de varianza constante, que se nota en el patrón "tipo corneta" que siguen los residuos en la medida de que el predicho crece. Sin embargo, la violación no es tan fuerte como para que pueda tener algún impacto en las conclusiones.

En la figura 6.23*a*) se grafican el orden de corrida contra los residuos, de ahí se observa que el supuesto de independencia se cumple, ya que los residuos caen alea-

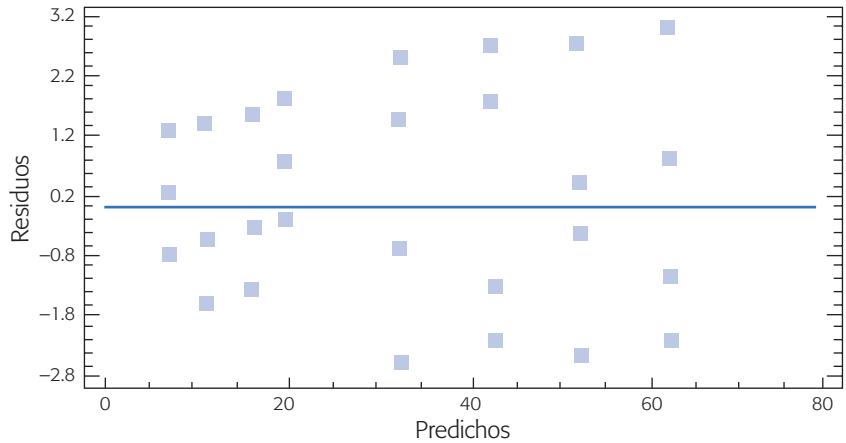


Figura 6.22 Residuos *vs.* predichos; ejemplo de semiconductores.

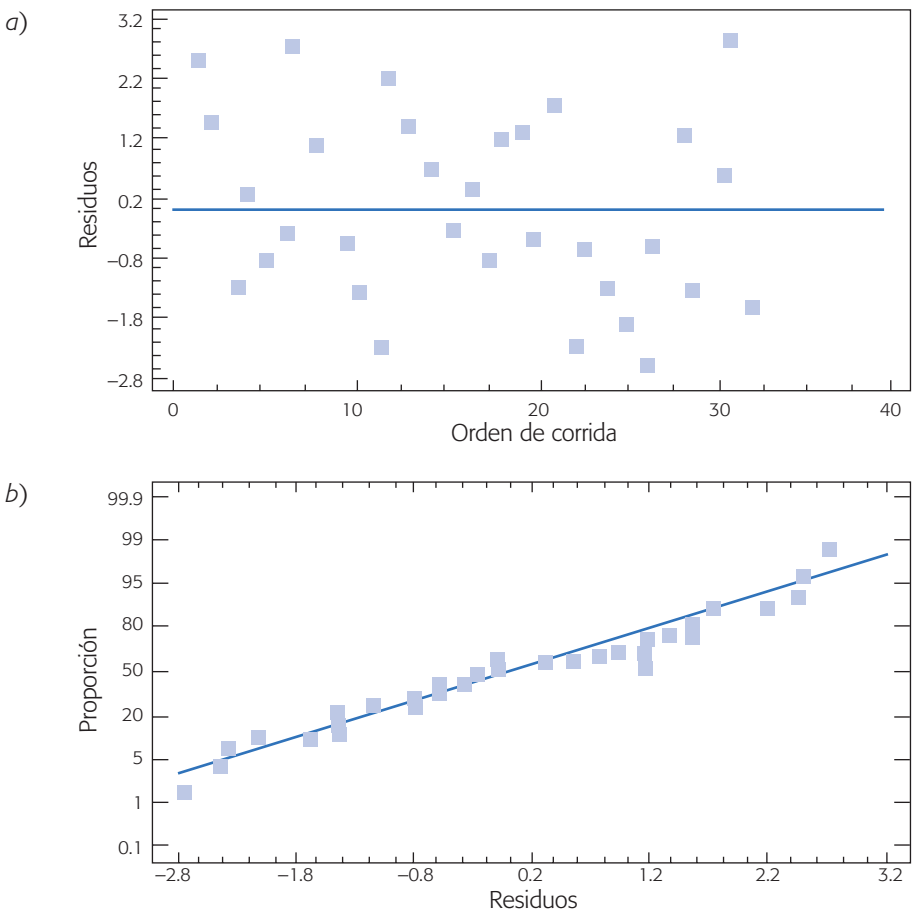


Figura 6.23 *a*) Residuos *vs.* orden de corrida; *b*) Residuos en papel normal, ejemplo de semiconductores.

toriamente en una banda horizontal. Asimismo, también se cumple el supuesto de normalidad de los residuos, ya que éstos se ajustan bien a una recta en el papel de probabilidad normal [véase figura 6.23*b*)].

Análisis alternativo: colapsación o proyección del diseño

En el análisis del experimento sobre rendimiento de semiconductores, una de la conclusiones es que no tuvieron ningún efecto los factores D (dimensión de la máscara) y E (tiempo de grabado). Este hecho da pie a *colapsar el diseño* en esas dos direcciones para convertirlo en un diseño factorial 2^3 con cuatro réplicas. Estas réplicas son más que suficientes para obtener un buen estimador del cuadrado medio del error en el ANOVA. La ganancia de la operación *colapsar* es la simplificación del análisis, al evitar la presencia en los resultados de los dos factores que ya se sabe no tienen ningún efecto significativo; además de que ahora se tienen repeticiones que permiten estimar directamente un buen cuadrado medio del error. En la figura 6.24 se dibuja el diseño original y el resultado de colapsarlo en dos direcciones. Observe las cuatro réplicas del diseño resultante.

Varios software tienen la opción de colapsar, entre ellos el *Statgraphics*, esto hace que no se requiera volver a capturar los datos. Si se quisieran los datos del diseño colapsado, la matriz de diseño se obtiene al ignorar las columnas de los factores que se quieren colapsar; en la notación de Yates se ignoran las letras que corresponden a dichos factores y surgen por sí solas las réplicas de cada tratamiento. Por ejemplo, en algún lugar de la tabla 6.14 se observaron los resultados:

$$a = 9,\ ad = 10,\ ae = 12,\ ade = 10$$

y al colapsar los factores D y E se eliminan las letras minúsculas correspondientes y estos resultados se convierten en:

$$a = 9,\ a = 10,\ a = 12,\ a = 10$$

que son las cuatro réplicas en el tratamiento $a \rightarrow (1, -1, -1)$ del diseño 2^3 resultante. Algo similar pasa con cada tratamiento.

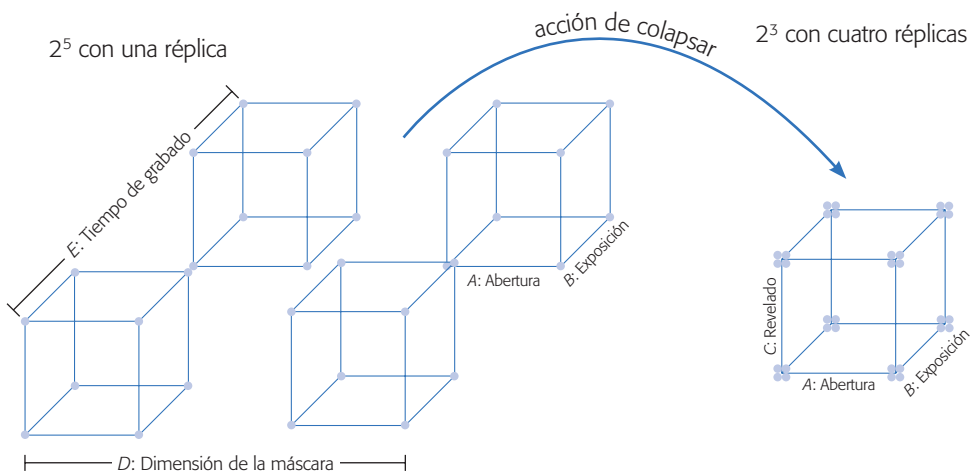


Figura 6.24 Colapsación en dos direcciones; ejemplo de semiconductores.

Tabla 6.18 ANOVA directo del diseño colapsado para semiconductores.

FV	SC	GL	CM	F_0	Valor-*p*
A: Abertu	1 116.28	1	1 116.28	361.11	0.0000
B: T-expo	9 214.03	1	9 214.03	2 980.68	0.0000
C: T-reve	750.78	1	750.78	242.87	0.0000
AB	504.03	1	504.03	163.05	0.0000
AC	1.53	1	1.53	0.50	0.4881
BC	0.03	1	0.03	0.01	0.9207
Error	77.28	25	3.09		
Total	11 664.0	31			

El análisis de varianza directo del diseño colapsado (excluyendo sólo a la interacción triple), se muestra en la tabla 6.18. Nótese que este análisis está más cerca del mejor ANOVA dado en la tabla 6.17; de aquí sólo resta eliminar a los efectos *AC* y *BC* para llegar al mejor análisis de varianza.

Omitimos la interpretación de los efectos activos, así como la verificación de supuestos del modelo porque se obtienen los mismos resultados que con el análisis sin colapsar, pero se deja como ejercicio al lector.

Cuando la significancia de los efectos es menos clara: un ejemplo

A continuación se muestra un experimento 2^5 no replicado, cuyo análisis es más "complicado" que el del ejemplo anterior. El objetivo es ilustrar las técnicas de separación de efectos en una situación menos ideal. El objetivo general del experimento es mejorar el entendimiento de cómo es que afectan el rendimiento los factores que típicamente se controlan en el proceso de fermentación de agave, en la elaboración del tequila.

Los factores que se decide estudiar y sus niveles en unidades originales son los siguientes:

A: Temperatura de carga (28°C, 30°C)
B: Inóculo de levadura (0.31%, 0.72%)
C: Temperatura medio ambiente (22°C, 34°C)
D: Temperatura de fermentación (34°C, 42°C)
E: Brix de carga (13.0°*BX*, 15.5°*BX*)

Se utiliza un diseño factorial 2^5 con sólo una corrida en cada tratamiento. El diseño y los rendimientos observados se muestran en la tabla 6.19.

Análisis del experimento

Efectos estimados y diagrama de Pareto. En este diseño el contraste de cada efecto tiene 32 términos, así que los cálculos a mano son engorrosos y es mejor

Tabla 6.19 Diseño factorial 2^5 con sólo una corrida por tratamiento, ejemplo del tequila.

Notación de Yates	A	B	C	D	E	Rend. (%)	Notación de Yates	A	B	C	D	E	Rend. (%)
(1)	–	–	–	–	–	60.30	*e*	–	–	–	–	+	79.44
a	+	–	–	–	–	73.20	*ae*	+	–	–	–	+	60.40
b	–	+	–	–	–	95.62	*be*	–	+	–	–	+	46.98
ab	+	+	–	–	–	79.41	*abe*	+	+	–	–	+	61.95
c	–	–	+	–	–	95.62	*ce*	–	–	+	–	+	57.05
ac	+	–	+	–	–	97.24	*ace*	+	–	+	–	+	42.09
bc	–	+	+	–	–	76.17	*bce*	–	+	+	–	+	47.65
abc	+	+	+	–	–	57.05	*abce*	+	+	+	–	+	53.69
d	–	–	–	+	–	90.76	*de*	–	–	–	+	+	61.58
ad	+	–	–	+	–	46.96	*ade*	+	–	–	+	+	27.00
bd	–	+	–	+	–	45.30	*bde*	–	+	–	+	+	61.15
abd	+	+	–	+	–	42.09	*abde*	+	+	–	+	+	57.05
cd	–	–	+	+	–	89.13	*cde*	–	–	+	+	+	56.12
acd	+	–	+	+	–	71.31	*acde*	+	–	+	+	+	55.36
bcd	–	+	+	+	–	76.17	*bcde*	–	+	+	+	+	29.30
abcd	+	+	+	+	–	61.58	*abcde*	+	+	+	+	+	28.59

utilizar un paquete estadístico. Por poner un ejemplo, el contraste del efecto *A* está dado por:

$$\begin{aligned}\text{Contraste } A &= a + ab + ac + abc + ad + abd + acd + abdc + ae + abe + ace + \\ &\quad abce + ade + abde + acde + abcde - (1) - b - c - bc - d - bd - cd \\ &\quad - bcd - e - be - ce - bce - de - bde - cde - bcde \\ &= 73.2 + 79.41 + 97.24 + 57.05 + 46.96 + 42.09 + 71.31 + 61.58 + \\ &\quad 60.4 + 61.95 + 42.09 + 53.69 + 27.0 + 57.05 + 55.36 + 28.59 - \\ &\quad 60.3 - 95.62 - 95.62 - 76.17 - 90.76 - 45.3 - 89.13 - 76.17 - \\ &\quad 79.44 - 46.98 - 57.05 - 47.65 - 61.58 - 61.15 - 56.12 - 29.3 \\ &= -153.37\end{aligned}$$

De acuerdo con la fórmula (6.14), el contraste se multiplica por 1/16 para obtener el efecto correspondiente, es decir,

$$\text{Efecto } A = \frac{1}{16}[\text{Contraste } A] = \frac{-153.37}{16} = -9.58563$$

Al hacer los cálculos para todos los efectos principales e interacciones de dos factores se obtienen los valores dados en la tabla 6.20. Se observa que los efectos que impactan más al rendimiento son: *E*, *D* y *CE*, pero falta ver si éstos son importantes estadísticamente.

En la figura 6.25 se muestra un diagrama de Pareto estandarizado, que incluye interacciones triples que siempre deben ser evaluadas. Como se aprecia, no es claro cuáles son los efectos significativos.

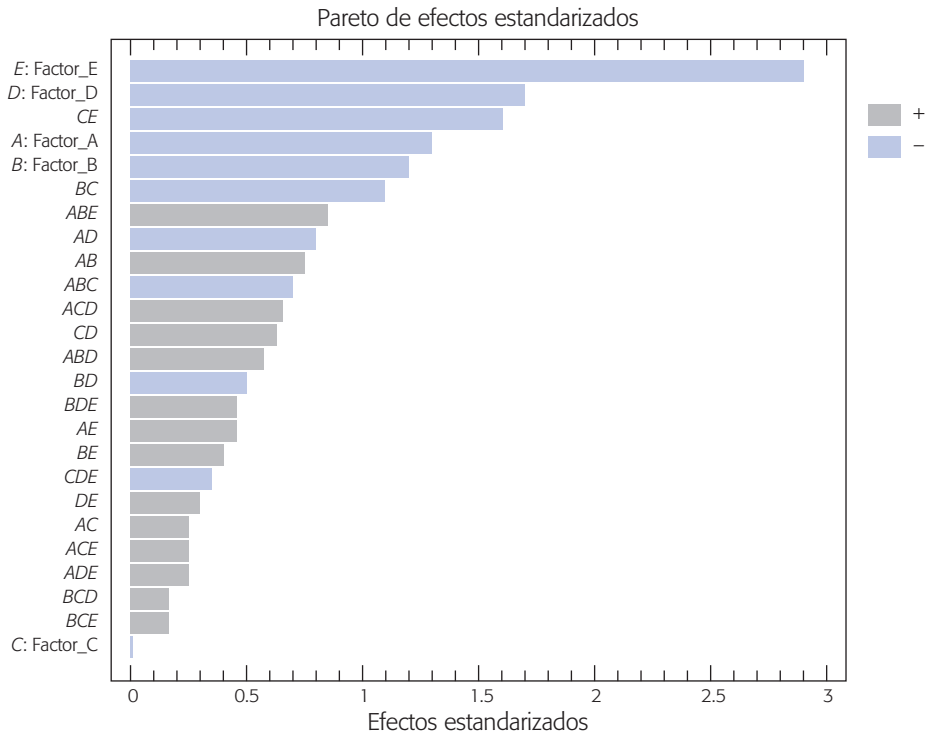


Figura 6.25 Diagrama de Pareto de efectos para el ejemplo del tequila.

Los escalones donde cambia bastante la longitud de las barras del Pareto separan grupos de efectos con diferente importancia. Así, es posible que sólo el efecto *E* sea significativo, ya que su barra es un poco más grande que las demás, pero también puede ocurrir que los primeros seis efectos sean significativos, ya que en este punto se encuentran el segundo escalón en tamaño.

Gráfica de efectos en papel normal. En la figura 6.26 se grafican los 31 efectos en *papel de probabilidad medio normal* (*half normal plot*); se observan seis efectos alejados de la línea, y son los mismos que el Pareto detecta en las primeras seis barras. Así, en apariencia, los efectos *A*, *B*, *D*, *E*, *BC* y *CE* son significativos. Para

Tabla 6.20 Efectos estimados.

media = 61.9784
A: Temp. de carga = –9.58563
B: Inóculo = –8.98812
C: Temp. medio ambiente = 0.308125
D: Temp. de fermentación = –11.5256
E: Brix de carga = –20.7819
AB = 5.96938; *AC* = 2.04812
AD = –5.36062; *AE* = 2.94313
BC = –7.72687; *BD* = –3.13563
BE = 2.40312; *CD* = 5.15062
CE = –11.0206; *DE* = 2.38812

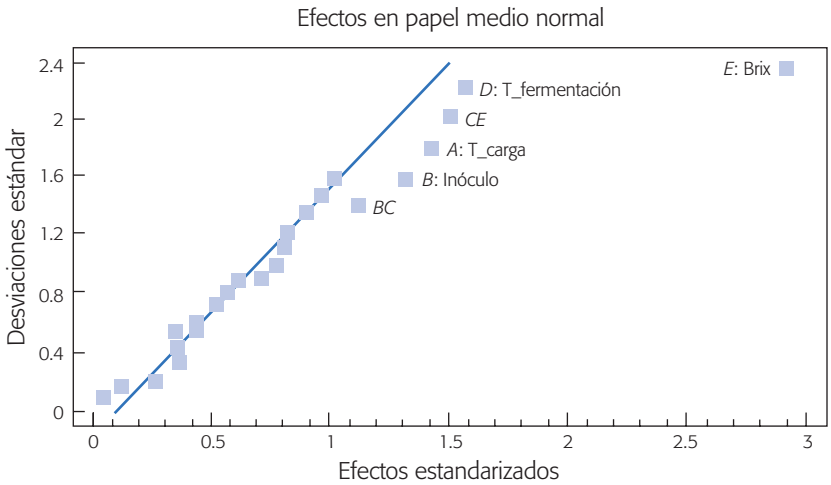


Figura 6.26 Efectos en papel normal para el ejemplo del tequila.

corroborarlo se decide construir un ANOVA (tabla 6.21), mandando al error a los restantes 25 efectos pequeños y aparentemente "no significativos".

Éste es el mejor ANOVA, de acuerdo con los criterios arriba señalados, de ahí se concluye que son significativos los efectos A, D, E y CE, a un nivel $\alpha = 0.05$. Los efectos B y BC no alcanzan a ser significativos a este nivel, pero no son del todo aleatorios, por eso no deben excluirse de la tabla. De hecho, el efecto B tiene un valor-$p = 0.064$ y también se recomienda considerarlo en la interpretación.

Verificación de supuestos

De acuerdo con las figuras 6.27*a* y *b*, se concluye que no hay problemas con los supuestos de varianza constante y de normalidad. En la primera, los puntos caen aleatorios sobre una banda horizontal, y en la segunda los residuos se ajustan bastante bien a una línea recta.

Gráficas de efectos y conclusiones

En la figura 6.28 se dibujan los efectos principales de A, B, D y E. Ahora interesa encontrar la combinación de niveles que da por resultado el mayor rendimiento. Se

Tabla 6.21 Análisis de varianza.

FV	SC	GL	CM	F_0	Valor-p
A: T-carga	735.074	1	735.074	5.27	0.0494
B: Inóculo	646.291	1	646.291	3.75	0.0641
D: T-fermen	1 062.72	1	1 062.72	6.17	0.0201
E: Brix	3 455.09	1	3 455.09	20.06	0.0001
BC	477.637	1	477.637	2.77	0.1083
CE	971.633	1	971.633	5.64	0.0255
Error	4 305.84	25	172.234		
Total		31			
$R^2_{aj} = 54.18$, $R^2 = 63.05$					

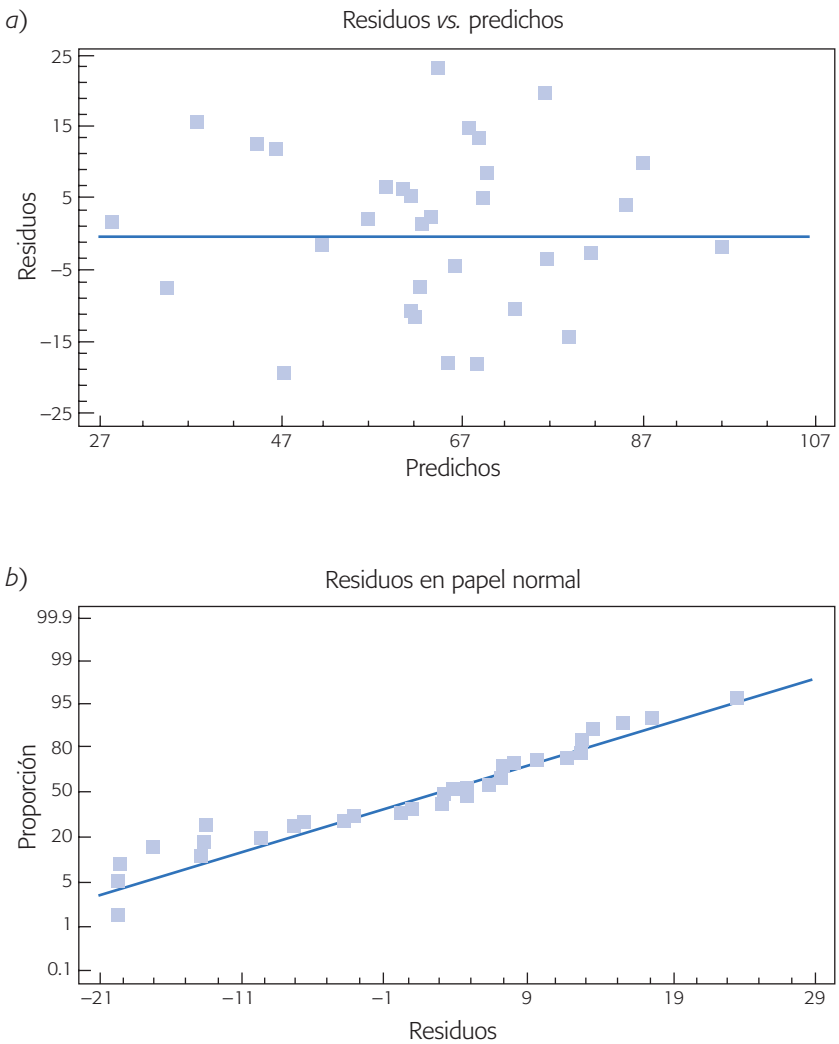


Figura 6.27 Gráfica de residuos para el ejemplo del tequila.

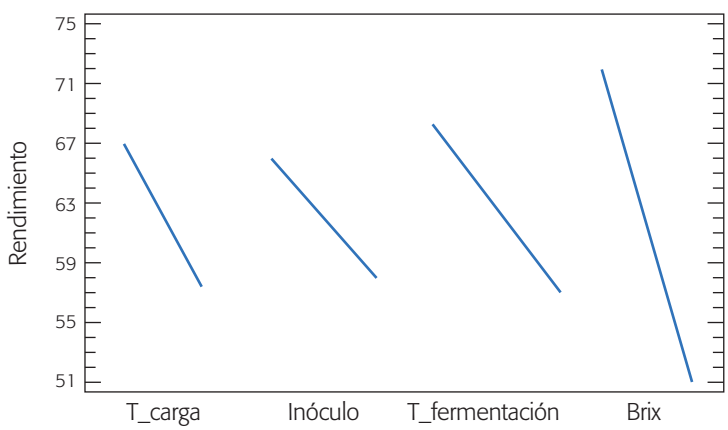


Figura 6.28 Efectos principales de *A, B, D* y *E*; ejemplo del tequila.

observa que conforme se aumenta el nivel de los factores A, B, D y E, el rendimiento disminuye; por lo tanto, los mejores niveles para estos factores son: A^-, B^- y D^-, respectivamente. Por la prioridad de las interacciones sobre los efectos principales, los mejores niveles de los dos factores restantes se obtienen de la gráfica de interacción.

El efecto de interacción CE se representa en la figura 6.29, en la cual se observa que el efecto de incrementar C es positivo (aumenta el rendimiento) siempre y cuando E se encuentre en su nivel menos (–); porque si E está en su nivel +, ocurre un efecto negativo. De esta forma, se busca el punto (extremo de línea) más alto con respecto a la escala del rendimiento, y se obtiene que los mejores niveles para los factores C y E son: C^+ y E^-. En resumen, el mejor tratamiento es cuando todos los factores, excepto C, se encuentran en su nivel bajo.

El efecto de interacción BC, que también se representa en la figura 6.29, no se interpreta por no tener la suficiente importancia. Sin embargo, aunque se interpretara, la conclusión anterior no cambia.

Predicción. El modelo lineal asociado al análisis de varianza de la tabla 6.21 explica, de acuerdo al coeficiente de determinación R^2_{aj}, 54.18% de la variabilidad en el rendimiento del proceso de fermentación del agave. Este valor tan bajo de R^2_{aj} no permite tener buena calidad en la predicción, ya que el exceso de variabilidad no explicada hace que el intervalo de confianza para el rendimiento futuro sea demasiado amplio. En la sección "Diseño factorial 2^2" de este capítulo, se estudian algunas de las causas por las que el coeficiente de determinación puede resultar bajo. De esa sección observamos que aun con R^2 bajos se pueden obtener conclusiones. En el caso del ejemplo tenemos conclusiones sólidas, y lo que seguiría es poner el proceso a correr en la mejor condición encontrada con el análisis, y ver cómo mejora el rendimiento. Lo que queda claro es que las predicciones con el modelo serán poco precisas, pero eso no quita que mejore el proceso en lo que se refiere a las condiciones en las que operaba antes del experimento. Al respecto y en el ejemplo, puede verificarse que el modelo ajustado predice un rendimiento de 96.79% en el mejor tratamiento, es decir,

$$\hat{Y}\,(-1, -1, +1, -1, -1) = 96.79$$

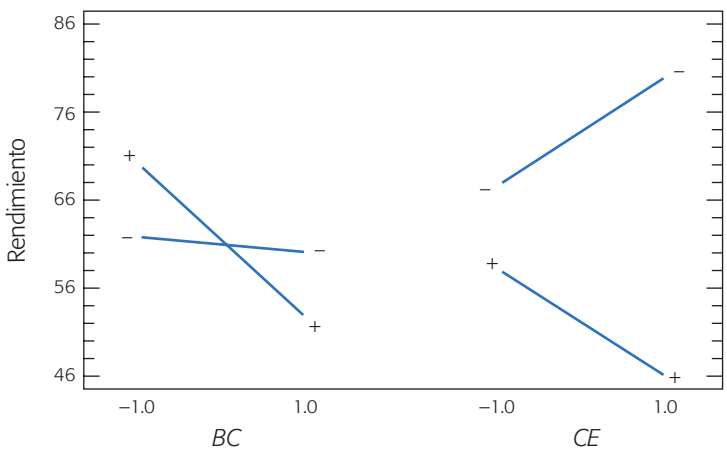


Figura 6.29 Efectos de interacción BC y CE; ejemplo del tequila.

El intervalo de confianza al 95% para el rendimiento promedio sobre este punto, está dado por [84.15, 109.43]. Por supuesto, que de 100% nunca se va a pasar, y es factible observar rendimiento promedio bajos, como 84.15%. Sin embargo, se debe poner a trabajar el proceso en esa condición y ver cómo mejora la situación prevaleciente. Pero, si las mejoras aún no son suficientes, habrá que generar más conocimiento sobre el proceso a fin de generar y confirmar nuevas conjeturas sobre las causas de la variabilidad. Parte de las respuestas se pueden encontrar mediante una mayor experimentación y conocimiento sobre la física del proceso. Por ejemplo, es viable aplicar la metodología de superficie de respuesta (capítulo 12) para encontrar mejores niveles de los factores controlados.

Factoriales 2^k con punto al centro

Punto al centro
Tratamiento formado por la combinación del nivel intermedio o medio de todos los factores en un diseño 2^k.

Cuando en un diseño factorial 2^k los k factores admitan un nivel de prueba intermedio, es recomendable implementar un tratamiento adicional formado por la combinación del nivel intermedio o medio de todos los factores. Esta combinación se le conoce como *punto central*. Hay dos razones por las que es deseable correr el punto central con cierto número de réplicas. La primera es obtener grados de libertad adicionales para el error en la tabla de ANOVA, sin perjudicar el balance en la estimación ni los efectos de interés. Ya se ha mencionado la conveniencia de interpretar ANOVA con al menos 8 grados de libertad en el error, condición que a veces es difícil de cumplir, por ejemplo cuando por razones económicas el experimento se corre sin las réplicas suficientes. Un ejemplo es correr el factorial 2^3 sin réplicas: se tienen 7 grados de libertad totales, por lo que es imposible construir un análisis de varianza con grados de libertad suficientes para el error. Cuatro o cinco repeticiones al centro agregarían esa cantidad de grados de libertad para el error, además de proveer de un estimador puro (independiente de los efectos estimados) de la varianza en dicho punto. Es más factible obtener cuatro corridas en el centro, que repetir los tratamientos del experimento completo. La segunda razón, dirigida a factores cuantitativos, es que las repeticiones al centro permiten detectar la posible presencia de curvatura en al menos uno de los factores objeto de estudio. La curvatura a la que nos referimos son los efectos cuadráticos A^2, B^2, … (véase figura 7.1). Una vez detectados estos efectos, el experimento se aumenta[2] con más puntos experimentales para analizar dicha curvatura.

Ejemplo 6.1

Factorial 2^3 con repeticiones al centro. En un proceso de circuitos integrados (obleas) interesa minimizar la corriente de fuga, que se supone depende de la temperatura de quemado (A), tiempo de quemado (B) y porcentaje de nitrógeno (C). Para ello se decide correr un experimento factorial 2^3 con dos réplicas y cuatro repeticiones al centro. Los resultados obtenidos se muestran enseguida:

[2] Los detalles de cómo aumentar el experimento se presentan en el capítulo 12, en la discusión relativa al diseño central compuesto.

Temp.	Tiempo	% de N	Y = Corriente de fuga
–1	–1	–1	2.153, 1.843
+1	–1	–1	1.609, 2.018
–1	+1	–1	1.346, 1.766
+1	+1	–1	1.695, 2.051
–1	–1	+1	3.864, 5.041
+1	–1	+1	7.054, 5.574
–1	+1	+1	5.519, 4.181
+1	+1	+1	5.746, 6.088
0	0	0	2.490, 2.384
0	0	0	2.474, 1.778

El ANOVA para este experimento se muestra en la tabla 6.22*a*). Se aprecia claramente que hay un efecto dominante (*C*), y dos efectos que están en duda (*A* y *AC*), y el resto no son significativos. Además, los coeficientes R^2 tienen un valor aceptable (86% y 78%). Para llegar al mejor ANOVA se eliminan los efectos que claramente no son significativos (*B*, *AB*, *BC* y *ABC*) y se obtiene que los efectos que estaban en duda (*A* y *AC*) son importantes al 10% de significancia.

Pero dado que se corrió el punto al centro, hay oportunidad de verificar curvatura. En la tabla 6.22*b* se muestra el mejor ANOVA y además el error se parte en los componentes de falta de ajuste (*lack-of-fit*) y error puro, donde *error puro* se obtiene de las repeticiones en el centro y en los puntos factoriales (las fórmulas se pueden ver en el capítulo 11). La falta de ajuste resulta significativa (valor-p = 0.0003). Como en este experimento se hicieron réplicas, entonces la falta de ajuste incluye los puntos al centro y los puntos replicados. De aquí que en este caso, el que la falta de ajuste sea significativa, es un fuerte indicio de curvatura. Este indicio se comprueba observando la gráfica de residuos contra los niveles de los factores (figura 6.30), donde se aprecia que los residuos correspondientes a las observaciones en el centro del experimento caen bastante abajo con respecto a los residuos en los extremos, lo cual es un indicativo de la presencia de curvatura en cada factor. El modelo predice valores de la corriente de fuga mayores a los observados en el centro, por eso los residuos ahí son negativos.

TABLA 6.22*a* ANOVA para corriente de fuga de obleas.

FV	*SC*	*GL*	*CM*	F_0	Valor-p
A: Temp.	2.342	1	2.342	3.16	0.1010
B: Tiem.	0.0365	1	0.0365	0.05	0.8283
C: % de N	51.072	1	51.072	68.82	0.0000
AB	0.0214	1	0.0214	0.03	0.8678
AC	1.954	1	1.954	2.63	0.1306
BC	0.0366	1	0.0366	0.05	0.8278
ABC	0.4199	1	0.4199	0.57	0.4664
Error	8.9049	12	0.7421		
Total	64.789	19			
$R^2 = 86\%,\ R^2_{aj} = 78\%$					

Tabla 6.22*b* Mejor ANOVA y prueba de falta de ajuste para corriente de fuga.

FV	*SC*	*GL*	*CM*	F_0	Valor-*p*
A: Temp.	2.342	1	2.342	9.05	0.0088
C: % de N	51.072	1	51.072	197.25	0.0000
AC	1.954	1	1.954	7.55	0.0150
Falta de ajuste	5.535	1	5.535	21.38	0.0003
Error puro	3.884	15	0.259		
Total	64.789	19			

Como comentario adicional, note que el factor *B* no afecta de ninguna manera la respuesta *Y*, el experimento se puede colapsar en un 2^2 + centro con cuatro réplicas (véase ejercicio 24). Al analizarse el diseño colapsado se llegaría a este mismo ANOVA.

Así, como hay curvatura, es necesario aumentar el experimento con puntos adicionales para estudiar o estimar los efectos A^2 y C^2 y saber cuál de ellos o si ambos provocan la falta de ajuste. En este caso el modelo de regresión asociado al ANOVA de la tabla 6.22*b* es:

$$\hat{Y} = 3.33 + 0.38x_1 + 1.78x_3 + 0.35x_1x_3 \qquad \textbf{(6.17)}$$

el cual no se ajusta bien a la respuesta observada en el punto al centro, ya que le faltan los términos x_1^2 y/o x_3^2. Después de detectar la curvatura, lo que sigue es correr puntos experimentales adicionales que permitan ajustar un modelo con términos cuadráticos (véase el ejemplo de esta estrategia en el capítulo 12), y así modelar de manera adecuada a *Y* en función de x_1 y x_3.

Factoriales 2^k en bloques

Por lo general, no es posible correr todos los tratamientos de un diseño factorial 2^k bajo las mismas condiciones experimentales, es decir, durante la planeación del

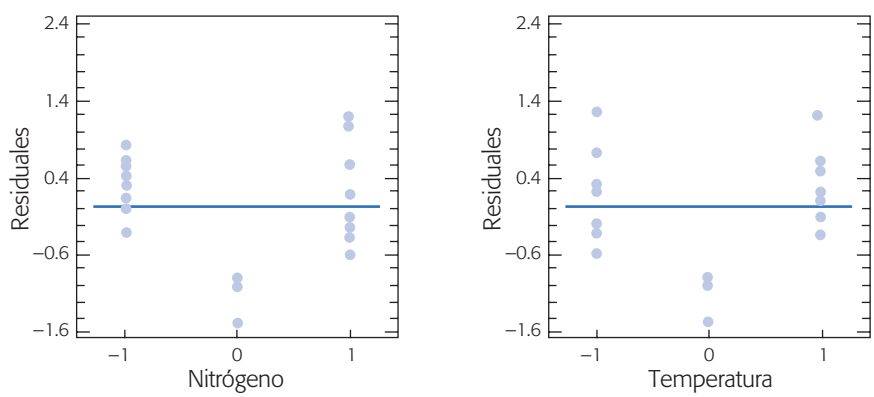


Figura 6.30 Residuos contra niveles de factores; ejemplo de corriente de fuga.

experimento aparece alguna restricción adicional que hace necesario considerar al menos un factor de bloque en el estudio (véase capítulo 4). El objetivo del experimento repartido en bloques es estudiar el efecto de los k factores sobre la respuesta y conocer la pertinencia de haberlo considerado. El uso adecuado de bloques incrementa la precisión del experimento, además es un medio para darle un rango de validez mayor al estudio: se tienen conclusiones válidas dentro de distintos bloques (días, máquinas, material, tipo de producto, etc.), que son inevitables y siempre están presentes en el proceso. Algunos casos típicos son los siguientes:

1. No es posible correr el factorial completo 2^k en el mismo día, ya sea porque las corridas o el proceso de medición son lentos, o por la cantidad de corridas. Si se considera que el factor día afecta los resultados del estudio, entonces se decide incorporarlo al experimento como factor de bloque. Si el proceso estudiado es sensible a los cambios de turno dentro del mismo día, los turnos deben considerarse como el factor de bloque.
2. Cuando un lote de material no alcanza para hacer todas las corridas experimentales y se sospecha que las diferencias entre lotes podrían sesgar los resultados, es necesario repartir de manera adecuada las corridas experimentales en varios lotes (bloques); o bien, cuando no se toma en cuenta que el material de prueba tiene dos o más características distintivas que pueden sesgar las conclusiones del estudio.
3. Cuando no es posible contar durante el experimento completo 2^k con el mismo operador o con el mismo instrumento de medición, y se sospecha que éstos pueden influir en el desempeño del proceso, entonces hay que considerarlos como factores de bloque.

En el experimento, no se pretende bloquear activamente todas las posibles fuentes de variación, sólo aquellas que puedan tener algún impacto importante sobre la respuesta de interés. En algunos casos bastará con mantener fijos los posibles factores de bloque durante las corridas experimentales. La estrategia para correr un diseño factorial 2^k en b bloques es distribuir de manera adecuada los 2^k tratamientos en los b bloques. Esta distribución se hace bajo el principio de jerarquía de efectos: son más importantes los efectos principales, seguidos por las interacciones dobles y luego las triples, cuádruples, etc. Es el mismo principio en que se basa la construcción de fracciones (capítulo 8), de manera que generar un bloque adecuado es lo mismo que generar una fracción adecuada. Entonces, al repartir los tratamientos en bloques se busca detectar lo menos posible el estudio de los efectos principales e interacciones dobles. La mayoría del software especializado incluye la alternativa de correr un diseño 2^k en bloques. Un caso que no perjudica la estimación de ningún efecto es que cada bloque consista en una réplica completa del experimento.

Cuando los bloques son las réplicas. Considere otra vez el factorial con cuatro réplicas discutido al inicio de este capítulo (ejemplo de la ranuradora). Suponga que cada una de las cuatro réplicas se corrió en un día diferente. El experimento repartido en cuatro bloques (días) se muestra en la siguiente tabla.

broc	veloc	Bloque I	Bloque II	Bloque III	Bloque IV	Total
–1	–1	18.2	18.2	12.9	14.4	64.4 = (1)
+1	–1	27.2	24.0	22.4	22.5	96.1 = *a*
–1	+1	15.9	14.5	15.1	14.2	59.7 = *b*
+1	+1	41.0	43.9	36.3	39.9	161.1 = *ab*
	Totales:	$Y_{\bullet\bullet 1} = 102.3$	$Y_{\bullet\bullet 2} = 101.3$	$Y_{\bullet\bullet 3} = 86.7$	$Y_{\bullet\bullet 4} = 91.0$	$Y_{\bullet\bullet\bullet} = 381.3$

De los totales por bloque, dados en la parte inferior de la tabla, se dice que la suma de cuadrados para los bloques es:

$$SC_{bloq} = \sum_{k=1}^{4} \frac{Y_{\bullet\bullet k}^2}{2\times 2} - \frac{Y_{\bullet\bullet k}^2}{N} = \frac{1}{4}(102.3^2 + 101.3^2 + 86.7^2 + 91.0^2) - \frac{381.3^2}{16} = 44.36$$

Y tiene 3 grados de libertad porque son cuatro bloques o días. A partir de ésta se incluye un renglón en el ANOVA donde se prueba la hipótesis H_0 : *Efecto de bloque* = 0 (tabla 6.23). Como el valor-*p* = 0.0280, se concluye que el efecto de bloque (debido a los días) es significativo. En la tabla de datos, observe que en los días 3 y 4 hubo menos vibración que en los primeros dos días. Sin embargo, sabemos que el orden de las corridas fue completamente aleatorio y en un mismo día; así que este efecto de bloque no es tal y ocurre sólo por azar. Si el efecto fuera real, se interpreta con gráficas de medias y pruebas de comparaciones múltiples a fin de comprender mejor la manera en que los días afectan la respuesta.

Las conclusiones sobre el mejor tratamiento serían las mismas que se obtuvieron anteriormente; en este caso, el efecto de bloque significativo implicaría que la respuesta esperada sobre el mejor tratamiento cambia de manera significativa día con día.

Generando los bloques con contrastes. Si los bloques se generan con los contrastes de alguna(s) interacción(es), entonces los efectos correspondientes y sus productos aparecerán confundidos con los bloques. La confusión consiste en que al estimar el efecto de bloque también se está calculando la interacción o interacciones utilizadas en su construcción. Mientras los efectos utilizados en la generación de los bloques sean de interacciones triples en adelante, el efecto confundido se puede atribuir sin mayor trámite a los bloques.

Tabla 6.23 ANOVA con efecto de bloque.

FV	SC	GL	CM	F_0	Valor-*p*
A: Broca	1 107.22	1	1 107.22	364.21	0.0000
B: Velocidad	227.25	1	227.25	74.75	0.0000
AB	303.63	1	303.63	99.88	0.0000
Bloque (día)	**44.36**	**3**	**14.79**	**4.86**	**0.0280**
Error	27.36	9	3.04		
Total	1 709.83	15			

Supongamos que se quiere correr un experimento 2^3 repartido en dos bloques, que se definen de acuerdo a los signos en el contraste de la interacción triple *ABC*. En las tablas que se presentan más adelante, los signos negativos de la columna *ABC* señalan el bloque 1 y los signos positivos el bloque 2.

Al estimar los efectos, el de interacción triple (efecto generador) estará confundido con el efecto de bloque, es decir, efecto *ABC* = *efecto de bloque*. El efecto observado se atribuye al bloque porque es más probable que éste sea lo que influye. Además, se sabe que las interacciones de tres factores en adelante generalmente no son significativas.

A	*B*	*C*	*AB*	*AC*	*ABC*
–	–	–	+	+	–
+	–	–	–	–	+
–	+	–	–	+	+
+	+	–	+	–	–
–	–	+	+	–	+
+	–	+	–	+	–
–	+	+	–	–	–
+	+	+	+	+	+

⇒

Corrida en bloque 1		
A	*B*	*C*
–	–	–
+	+	–
+	–	+
–	+	+

y

Corrida en bloque 2		
A	*B*	*C*
+	–	–
–	+	–
–	–	+
+	+	+

En general, el factorial 2^k se puede partir en 2^b bloques, para lo cual se requieren b efectos generadores iniciales, cuyos signos definen las combinaciones que componen a cada bloque. Los generadores iniciales se seleccionan de las interacciones de mayor orden, de manera que todos sus posibles productos también sean interacciones del más alto orden. Tanto los efectos iniciales como sus posibles productos estarán confundidos con bloques, es decir, no se podrán estudiar. En la tabla 6.24 se proveen generadores iniciales adecuados para construir 2^b bloques de un factorial completo 2^k, para algunos valores de k. También se puede usar un software estadístico para generar los bloques deseados para cualquier factorial completo o fraccionado.

Tabla 6.24 Efectos adecuados para generar bloques.

Número de factores (k)	Número de bloques (2^b)	Tamaño de bloque (2^{k-b})	Efectos generadores	Efectos confundidos con bloques
3	2	4	*ABC*	*ABC*
4	2	8	*ABCD*	*ABCD*
4	4	4	*ABC*, *ACD*	*ABC*, *ACD*, *BD*
5	2	16	*ABCDE*	*ABCDE*
5	4	8	*ABC*, *CDE*	*ABC*, *CDE*, *ABDE*
5	8	4	*ABE*, *BCE*, *CDE*	*ABE*, *BCE*, *CDE*, *AC* *ABCD*, *BD*, *ADE*
6	2	32	*ABCDEF*	*ABCDEF*
6	4	16	*ABCF*, *CDEF*	*ABCF*, *CDEF*, *ABDE*
6	8	8	*ABEF*, *ABCD*, *ACE*	*ABEF*, *ABCD*, *ACE*, *CDEF*, *BCF*, *BDE*, *ADF*

Uso de software estadístico

En *Statgraphics*, la secuencia a seguir para generar un diseño factorial 2^k es la siguiente: *Special* → *Experimental* → *Design* → *Create Design*; a partir de ahí, en *Design Class* se elige *Screening*, se da el número de variables de respuesta y el número de factores. Enseguida se puede introducir el nombre y los niveles de los factores, después se elige el diseño específico y el número de réplicas. De la versión 15 en delante de este software la secuencia es más directa y empieza por *DOE* → *Design Creation*.

Al momento de seleccionar el diseño, el primero del menú de opciones es el diseño factorial completo. Se le llama diseño base porque todavía falta por decidir cuántas réplicas han de hacerse y si se desean repeticiones al centro para llegar al arreglo definitivo, el cual se puede pedir en orden estándar o en orden aleatorio. Si se trata de resolver un ejercicio del libro, cuyos datos están acomodados en el orden estándar, es mejor desactivar la aleatorización para capturar los datos cómodamente; si se trata de la planeación de un nuevo experimento real, se pide en orden aleatorio (*randomnize*). En el menú de opciones también aparecen los diseños factoriales fraccionados del capítulo 8, así como los factoriales completos 2^k en bloques.

La impresión en papel de la hoja de trabajo (*worksheet*) que se encuentra en la opción tabular permite tener por escrito las corridas experimentales en orden aleatorio, y deja un espacio en blanco para poner el resultado de la respuesta que se obtiene al correr el experimento en cada combinación. El mismo arreglo de esta hoja se puede salvar o guardar como un archivo de diseño generado por *Statgraphics*. Una vez que se registran los resultados obtenidos se analiza el experimento. Para ello se sigue la misma secuencia, pero en lugar de elegir *Create Design*, se selecciona *Analyze Design*.

A continuación se da la variable de respuesta a analizar. Si hay más de una variable de respuesta se hace un análisis por separado para cada variable y al final se toman soluciones que beneficien lo más posible a todas las variables de respuesta (véase capítulo 13). Cuando el experimento tiene más de una réplica conviene verificar la pertinencia de los bloques: de manera automática, el software considera que cada réplica del experimento se correrá en circunstancias experimentales distintas, es decir, cada una posiblemente con diferentes lotes, en distintos días, turnos, etc. Si éste fuera el caso se analizan dichos bloques y se atribuyen al factor utilizado para formarlos. Si el experimento se corre completo bajo las mismas circunstancias (en un solo bloque) y que por lo tanto no puede existir este tipo de efectos, se activa la opción de análisis que ignora los bloques. En las opciones tabulares y gráficas se consideran los distintos procedimientos que se estudiaron en el presente capítulo.

En Minitab. La secuencia de opciones para crear un diseño factorial 2^k es la siguiente: *Stat* → *DOE* → *Factorial* → *Create Factorial Design*, donde se elige la opción de 2-*Level Factorial* y el número de factores. En la pestaña *Designs* aparecen los posibles diseños, el primero de los cuales es el completo. Abajo se define el número de puntos centrales, el número de réplicas y el número de bloques. En la pestaña *Factors* se puede dar el nombre y los niveles de los factores, y en *Options* se decide si los tratamientos se ordenan en forma aleatoria (*Randomize runs*). El análisis se hace con

la misma secuencia, pero en lugar de *Create* se elige *Analyze Factorial Design* y/o *Factorial Plots* (Gráficas).

En Excel. El análisis de los diseños factoriales en Excel, con excepción del factorial 2^2, debe hacerse "manualmente" porque no existen otros procedimientos. El factorial 2^2 se analiza por medio de la secuencia: *Herramientas* → *Análisis de datos* → *Análisis de varianza de dos factores con varias muestras por grupo*. Por ejemplo, para analizar con este procedimiento los datos del ejemplo de la ranuradora, éstos se acomodan en la hoja de Excel de la siguiente manera:

	B1	B2
V1	18.2	27.2
	18.9	24.0
	12.9	22.4
	14.4	22.5
V2	15.9	41.0
	14.5	43.9
	15.1	36.3
	14.2	39.9

En el rango de entrada se declara toda la matriz incluyendo las columnas con los rótulos y se le indican cuatro filas por muestra. En el ANOVA resultante el factor columna es la broca y el factor muestra es la velocidad. Los efectos se pueden obtener a partir de las sumas de cuadrados correspondientes.

Preguntas y ejercicios

1. Conteste las siguientes preguntas, de índole general, con respecto a los diseños factoriales:
   a) Explique qué son los factores y qué es la(s) variable(s) de respuesta.
   b) ¿Cuál es el objetivo de un diseño factorial?
   c) Ejemplifique y explique en qué consiste la estrategia de modificar o mover un factor a la vez, que es propia de la experimentación empírica.
   d) Señale y argumente qué ventajas tienen los experimentos factoriales sobre la estrategia de mover un factor a la vez.
   e) ¿Qué significa que un factor tenga un efecto significativo? ¿Sobre quién es el efecto?
   f) ¿Todos los factores deben ser de tipo cuantitativo o es posible involucrar factores cualitativos, por ejemplo dos tipos de máquinas, o la presencia o ausencia de alguna sustancia?
2. Suponga un diseño factoria1 2^2, cuyos factores y niveles son: temperatura (8, 20) y velocidad (4, 7). La variable de respuesta es rendimiento. Conteste las siguientes preguntas:
   a) ¿Por qué este diseño recibe tal nombre?
   b) Anote los diferentes tratamientos que forman este diseño. Utilice diferentes tipos de códigos.
   c) Represente en forma geométrica al diseño y resalte la región de experimentación.
   d) Explique cómo piensa que fue el proceso para seleccionar esos factores y esos niveles.
   e) Defina qué son los efectos principales y cuál el efecto de interacción.

f) Señale los diferentes efectos que se pueden estudiar con este diseño y la forma en que se calcula cada uno.
g) Describa en qué consiste la aplicación de los tres principios básicos del diseño de experimentos (capítulo 1), en este caso.

3. A continuación se muestran los resultados de un diseño factorial. Conteste los siguientes incisos sin utilizar un software computacional, es decir, haga las operaciones de manera manual.

		Réplica			
A	*B*	I	II	III	Total
–	–	82	80	84	(1) = 246
+	–	78	82	79	(*a*) = 239
–	+	71	70	66	(*b*) = 207
+	+	89	88	93	(*ab*) = 270

a) ¿Qué nombre recibe este diseño y por qué?
b) ¿Cuántos tratamientos tiene este diseño, cuántas réplicas?
c) En total son 12 corridas experimentales. Señale en qué orden debieron correrse y explique por qué.
d) Explique los efectos que se pueden estudiar a través de este diseño.
e) Obtenga los contrastes para los efectos principales de *A* y *B*, así como para la interacción.
f) Calcule los efectos principales y el efecto de interacción.
g) Haga las gráficas de los efectos principales de *A* y *B*, e interprételas.
h) Realice la gráfica de la interacción entre los factores *A* y *B*, e interprétela con detalle.
i) ¿Desde su punto de vista el factor *B* parece tener influencia sobre *Y*? Argumente su respuesta.

4. Suponga un diseño factorial 2^3, y conteste las siguientes preguntas.
a) Utilice la notación de (–, +) para los niveles de los factores, y escriba todos los tratamientos que forman este diseño.
b) Represente en forma geométrica este diseño y resalte la región de experimentación.
c) ¿Cuáles son todos los posibles efectos que se pueden estudiar con este diseño?
d) Para cada uno de los efectos anteriores, obtenga su contraste.
e) Señale en forma específica cómo utilizaría los contrastes para calcular los efectos y la suma de cuadrados.
f) En este caso, ¿cómo aplicaría los tres principios básicos del diseño de experimentos (capítulo 1)?

5. A continuación se muestran los resultados obtenidos en un diseño factorial 2^3 no replicado. Conteste los siguientes incisos sin utilizar un software computacional, es decir, haga las operaciones de manera manual.

¿Código?	*A*	*B*	*C*	*Y*
	–	+	–	25
	+	+	+	12
	–	–	–	30
	+	–	+	10
	–	–	+	10
	+	+	–	14
	–	+	+	31
	+	–	–	17

a) En la primera columna de la matriz de diseño especifique el código de cada uno de los tratamientos, de acuerdo a la notación de Yates.
b) Calcule los efectos principales de *A* y *B*.
c) Haga la gráfica de los efectos principales de *A* y *B*, e interprétela.
d) Calcule el efecto de la interacción de *AB*.
e) Realice la gráfica de la interacción entre los factores *A* y *B*, e interprétela con detalle.
f) ¿Qué tendría que hacer para saber si los efectos que calculó en los incisos anteriores afectan de manera significativa la variable de respuesta?
g) Calcule la suma de cuadrados para el efecto principal de *A* y para la interacción.

6. Suponga un diseño factorial 2^4, y conteste las siguientes preguntas.
a) Anote la matriz de diseño, es decir, haga una lista de todos los tratamientos que forman este diseño.
b) ¿Por qué este diseño recibe tal nombre?
c) ¿Cuáles son todos los posibles efectos que se pueden estudiar con este diseño?
d) Con respecto al análisis, ¿en qué consiste y cuál es el objetivo de obtener el mejor ANOVA?
e) ¿Cómo se calculan los coeficientes de determinación R^2 y R^2_{AjS}?
f) Si después de conseguir el mejor ANOVA, se obtiene que estos R^2_{aj} coeficientes tienen un valor de alrededor de 90, ¿qué significa esto?
g) Si por el contrario, tales coeficientes tienen un valor de alrededor de 20, ¿qué significa esto?
h) Obtenga el contraste para el efecto principal de *D* y para el efecto de interacción *CD*.
i) Señale en forma específica cómo utilizaría los contrastes para calcular los efectos y la suma de cuadrados.
j) ¿Puede darse el caso de que el efecto principal de *A* no sea significativo, y el efecto de la interacción *AB* sí lo sea?

7. En una fábrica de dientes se tiene problemas con la calidad: porosidad (burbujas de aire dentro de los dientes), manchas blancas, dientes sucios, dientes quebrados. En los intentos por resolver los problemas han hecho cambios en algunos factores o variables del proceso. Con base en la metodología del *DDE* se decide correr un diseño de experimentos 2^3. Los factores y niveles son: temperatura de prensado (90, 130°C), tiempo de prensado (8 y 15 minutos) y tamaño de partícula (sin tamizar y con tamizado), la variable de respuesta fue porcentaje de diente bueno en cada corrida (un lote de producción). Los datos son los siguientes:

Temperatura	Tiempo	T. de partícula	Porcentaje de diente bueno
90	8	Sin tamizar	76.4, 76.9
130	8	Sin tamizar	76.3, 76.9
90	15	Sin tamizar	80.4, 81.0
130	15	Sin tamizar	77.9, 79.6
90	8	Con tamizado	84.4, 84.6
130	8	Con tamizado	84.7, 84.5
90	15	Con tamizado	82.7, 83.2
130	15	Con tamizado	85.0, 84.7

a) Estime todos los posibles efectos y diga cuáles son significativos.
b) Realice un análisis de varianza de estos datos y obtenga conclusiones generales.
c) Verifique residuos, ¿qué observa que sea notorio?
d) ¿Hay un tratamiento ganador?
e) ¿Qué condición de proceso sugiere emplear a futuro? Tome en cuenta, además, que a mayor tiempo y mayor temperatura, más costos.
f) Las condiciones que se utilizaban antes del experimento eran: temperatura de 130°C y tiempo de 15 minutos. ¿Por qué cree que se eligieron niveles inferiores de prueba para estos factores?
g) Estos resultados, aunque positivos, no son suficientes; por lo tanto, qué sugiere usted, ¿explorar más niveles de los factores ya estudiados?, ¿considerar otras causas? o ¿qué? Argumente.

8. En una empresa lechera se han tenido problemas con la viscosidad de cierta bebida de chocolate. Se cree que con tres ingredientes que se agregan en pequeñas cantidades se puede resolver este problema, por lo que es necesario explorar la situación; para ello se corre un experimento 2^3 con dos réplicas. A continuación se aprecian los resultados obtenidos:

Ingrediente *A*	Ingrediente *B*	Ingrediente *C*	Viscosidad
–1	–1	–1	13.3, 13.9
+1	–1	–1	14.7, 14.4
–1	+1	–1	14.6, 14.9
+1	+1	–1	14.3, 14.1
–1	–1	+1	16.9, 17.2
+1	–1	+1	15.5, 15.1
–1	+1	+1	17.4, 17.1
+1	+1	+1	18.9, 19.2

a) Estime todos los posibles efectos y diga cuáles son significativos.
b) Realice un análisis de varianza de estos datos y obtenga conclusiones generales.
c) Interprete a detalle los efectos significativos.
d) ¿Hay un tratamiento ganador para minimizar?
e) Verifique residuos, ¿qué considera destacado?

9. En una empresa de electrónica una máquina toma componentes que le proporciona un alimentador, para montarlos o depositarlos en una tarjeta. Se ha tenido el problema de que la máquina falla en sus intentos por tomar el componente, lo cual causa paros de la máquina que detienen el proceso hasta que el operador se da cuenta y reinicia el proceso. Para diagnosticar mejor la situación, se decide correr un diseño de experimentos 2^4 con $n = 2$ réplicas, en el que se tienen los siguientes factores y niveles (–, +), respectivamente: *A*) Velocidad de cam (70%, 100%), *B*) Velocidad de mesa (media, alta), *C*) Orden o secuencia de colocación (continua, variable), *D*) Alimentador (1, 2). Como el proceso es muy rápido, es necesario dejarlo operar en cada condición experimental el tiempo suficiente para reproducir el problema. Se consideró que esto se lograba con suficiente confianza con 500 componentes; por ello, cada una de las corridas experimentales consistió en colocar 500 componentes, y se midieron dos variables de respuesta: $Y1$ = número de errores (o intentos fallidos), y $Y2$ = tiempo real (en segundos)

para tomar y "colocar" los 500 componentes. Es evidente que se quiere minimizar ambas variables. Los datos obtenidos se muestran en la siguiente tabla.

				Réplica 1		Réplica 2	
Factor *A*	Factor *B*	Factor *C*	Factor *D*	*Y*1	*Y*2	*Y*1	*Y*2
–1	–1	–1	–1	61	88	50	79
+1	–1	–1	–1	105	78	98	74
–1	+1	–1	–1	61	82	40	82
+1	+1	–1	–1	104	73	145	79
–1	–1	+1	–1	0	88	35	100
+1	–1	+1	–1	35	84	22	82
–1	+1	+1	–1	50	89	37	88
+1	+1	+1	–1	57	79	71	81
–1	–1	–1	+1	12	77	19	75
+1	–1	–1	+1	60	66	57	64
–1	+1	–1	+1	9	84	19	73
+1	+1	–1	+1	72	93	61	66
–1	–1	+1	+1	0	86	0	82
+1	–1	+1	+1	10	76	1	77
–1	+1	+1	+1	3	84	7	86
+1	+1	+1	+1	15	75	15	73

a) Al observar los datos obtenidos se deduce que hay algunos tratamientos que tienen pocos o ningún componente caídos, como por ejemplo el (–1, –1, +1, +1), alguien muy "práctico" decidiría poner la máquina a operar bajo estas condiciones, y olvidarse del análisis estadístico. De proceder así, explique qué información se perdería.
b) Investigue qué efectos influyen de manera significativa sobre *Y*1 (apóyese en Pareto y ANOVA).
c) Obtenga el mejor ANOVA.
d) Si en el análisis anterior encuentra alguna interacción significativa, analice con detalle la más importante e interprete en términos físicos.
e) ¿Qué tratamiento minimiza *Y*1?
f) Ahora investigue qué efectos influyen de manera relevante sobre *Y*2.
g) ¿Qué tratamiento minimiza *Y*2?
h) Encuentre una condición satisfactoria tanto para minimizar *Y*1 como *Y*2.
i) De los análisis de varianza para *Y*1 y *Y*2 observe el coeficiente R^2. ¿Qué concluye de ello?
j) Verifique residuos.

10. Un fabricante de bolsas de papel desea analizar la resistencia al rasgamiento (*Y*), para lo cual utiliza una escala numérica. Examina tres factores, cada uno en dos niveles, x_1 = papel, x_2 = humedad, x_3 = dirección del rasguño. Decide obtener tres observaciones (réplicas) en cada combinación, las mismas que se muestran en la siguiente tabla:

x_1	x_2	x_3	Resistencia		
–	–	–	3.8	3.1	2.2
+	–	–	6.6	8.0	6.8
–	+	–	3.4	1.7	3.8
+	+	–	6.8	8.2	6.0
–	–	+	2.3	3.1	0.7
+	–	+	4.7	3.5	4.4
–	+	+	2.1	1.1	3.6
+	+	+	4.2	4.7	2.9

a) Haga el análisis de varianza para estos datos.
b) Interprete los efectos significativos y encuentre el mejor tratamiento.
c) Verifique los supuestos del modelo.

11. En el área de SMT se busca reducir los defectos ocasionados por impresiones de soldadura en pasta inadecuada. Se corre un diseño 2^4 con dos réplicas y dos puntos centrales por réplica. Los factores son: altura de la mesa (*A*), velocidad de separación (*B*), velocidad de impresión (*C*) y presión de las escobillas (*D*). La variable de respuesta es la altura de la impresión de soldadura en pasta. El experimento se corrió en planta, pero como el proceso es muy rápido (la impresión de una tarjeta tarda menos de un minuto), entonces se recomienda obtener más de un producto en cada condición experimental. Por ello se decidió que cada prueba experimental debería de consistir en dejar que el proceso se estabilizara y a partir de ahí imprimir 10 tarjetas de manera consecutiva, a cada tarjeta se le midió la altura. Con estos 10 datos se calculó la media y la desviación estándar, para así analizar el efecto de los factores sobre ambas. Una vez que se corre en orden aleatorio la primera réplica de todos los tratamientos, se deja de experimentar y al día siguiente se hace de manera similar la segunda réplica. Los datos se muestran a continuación:

				Réplica 1		Réplica 2	
Factor *A*	Factor *B*	Factor *C*	Factor *D*	Media	D. estándar	Media	D. estándar
–1	–1	–1	–1	6.8	0.17	6.3	0.18
+1	–1	–1	–1	6.9	0.28	6.6	0.51
–1	+1	–1	–1	6.4	0.17	5.8	0.41
+1	+1	–1	–1	6.6	0.29	6.6	0.19
–1	–1	+1	–1	6.8	0.27	6.5	0.19
+1	–1	+1	–1	8.7	0.80	7.3	0.75
–1	+1	+1	–1	6.7	0.16	6.4	0.21
+1	+1	+1	–1	7.8	0.64	7.1	0.60
–1	–1	–1	+1	5.5	0.28	5.3	0.15
+1	–1	–1	+1	5.8	0.51	5.4	0.24
–1	+1	–1	+1	5.8	0.14	5.3	0.21
+1	+1	–1	+1	5.5	0.19	5.4	0.13
–1	–1	+1	+1	6.1	0.29	6.0	0.34
+1	–1	+1	+1	6.6	0.38	6.2	0.50
–1	+1	+1	+1	6.6	0.26	5.6	0.25
+1	+1	+1	+1	6.7	0.22	6.3	0.37
0	0	0	0	6.5	0.25	6.0	0.53
0	0	0	0	6.4	0.27	5.8	0.50

a) ¿Con qué finalidad se utilizan los puntos centrales?
b) Investigue qué efectos influyen de manera significativa sobre la altura promedio de la pasta (apóyese en Pareto y ANOVA).

c) Si en el análisis anterior encuentra alguna interacción significativa, analice a detalle la más importante.
d) Si se quiere un valor de 6.0 para la altura de la pasta, ¿cuáles son las condiciones para lograrlo?
e) Ahora investigue qué efectos influyen de manera relevante sobre la variabilidad de la altura de la pasta.
f) Encuentre una condición satisfactoria tanto para la altura como para minimizar la variabilidad.
g) De los análisis de varianza para la media y la desviación estándar vea el coeficiente R^2. ¿Qué concluye de ello?
h) ¿Hay evidencia de curvatura?
i) Verifique residuos.

12. En la refinación de pulpa de madera interesa estudiar cómo se afecta la calidad de la fibra al introducirle cargas de material inorgánico con dos diferentes métodos o procesos (mecánico e *in situ*), y ver cómo interactúa al considerar otros factores. Para ello se decide correr un diseño 2^4 no replicado, con los siguientes factores y niveles.

Factor	Niveles (bajo, alto)	
A: Proceso	Mecánico	*In situ*
B: Velocidad de agitación (rpm)	2 000	3 000
C: Tiempo (minutos)	30	60
D: Consistencia de la pulpa (%)	0.5	2.0

Se midieron cuatro variables de respuesta: Y_1: cenizas (%), Y_2: tensión (m), Y_3: blancura (%) y Y_4: opacidad (%). Los resultados para los 16 tratamientos se muestran en el siguiente cuadro, en el orden que se corrieron.

Tratamiento	Y_1	Y_2	Y_3	Y_4
c	0.48	579	86.17	77.85
bcd	1.46	692	86.5	76.82
bc	0.94	581	86.75	76.16
(1)	0.49	671	87.22	76.68
ad	5.5	653	88.55	79.3
b	1.49	867	87.62	77.15
a	7.49	496	88.77	79.75
ab	11.59	467	88.05	80.35
ac	13.23	437	87.3	78.72
bd	2.21	631	86.45	78.03
cd	5.06	565	85.75	79.42
abcd	7.78	549	88.45	80.81
abd	11.75	460	88.77	81.43
abc	12.57	462	89.42	81.55
d	0.72	620	87.67	77.87
acd	9.61	568	88.45	78.78

a) Bosqueje una gráfica de proceso en el que se muestren por un lado los factores controlados y por el otro las variables de respuesta.
b) Investigue qué efectos influyen de manera significativa sobre Y_1 (apóyese en Pareto y ANOVA).
c) Si se considera que el factor de mayor interés es el *A,* comente la forma en que actúa sobre Y_1.
d) Comente con detalle la manera en que interactúa el factor *A* con los otros factores, sobre Y_1.

e) ¿Qué tratamiento maximiza Y_1?
f) Verifique supuestos para la variable Y_1.
g) Repita los incisos anteriores para las otras variables de respuesta.
h) Encuentre condiciones satisfactorias para maximizar las cuatro variables de respuesta.
i) Haga un resumen del análisis realizado y destaque las principales conclusiones.

13. Se quiere aumentar el rendimiento de un proceso, y para ello se estudian tres factores con dos niveles cada uno. Se hacen tres repeticiones en cada tratamiento del diseño factorial 2^3 resultante. La variable de respuesta que se mide es rendimiento. Los datos son los siguientes:

Tratamiento	Repetición		
	1	2	3
(1)	22	31	25
a	32	43	29
b	35	34	50
ab	55	47	46
c	44	45	38
ac	40	37	36
bc	60	50	54
abc	39	41	47

a) ¿Cuáles efectos están activos?
b) Si obtuvo una interacción importante, interprétela con detalle.
c) Determine las condiciones de operación que maximizan el rendimiento.
d) ¿Cuál es la respuesta esperada en el mejor tratamiento?
e) Verifique los supuestos del modelo.

14. En una fábrica de componentes electrónicos, uno de los principales clientes reportó tener problemas con algunos de los productos (comportamiento eléctrico intermitente). Mediante el análisis de las muestras retornadas por el cliente, se identificó que el problema se relaciona con alambre mal colocado y podía obedecer a varias causas. Se decide correr una réplica de un experimento factorial 2^5, utilizando los siguientes factores y niveles:

Factor	Niveles (bajo, alto)	
A: Patrón de reconocimiento	un punto	dos puntos
B: Sistema de luz	fibra	incandescente
C: Umbral (*threshold*)	725	850
D: Colocación del dado	girado	normal
E: Brillo de la oblea	brillo	normal

La respuesta a medir es el número de unidades con alambre mal colocado. Cada prueba se hizo en la línea de ensamble y consistió en colocar cierta cantidad de alambres, que lo hace un equipo automático. La cantidad de alambres a colocar en cada prueba,

bajo cada tratamiento, se determinó de tal forma que tuviera alta probabilidad de detectar piezas con alambres mal colocados. Los datos son los siguientes:

$(1) = 105$	$d = 0$	$e = 34$	$de = 0$
$a = 0$	$ad = 0$	$ae = 3$	$ade = 0$
$b = 66$	$bd = 0$	$be = 18$	$bde = 0$
$ab = 7$	$abd = 5$	$abe = 2$	$abde = 0$
$c = 54$	$cd = 25$	$ce = 0$	$cde = 0$
$ac = 1$	$acd = 1$	$ace = 0$	$acde = 0$
$bc = 41$	$bcd = 0$	$bce = 49$	$bcde = 0$
$abc = 0$	$abcd = 0$	$abce = 4$	$abcde = 0$

a) Dibuje el diagrama de Pareto y el gráfico de Daniel considerando todas las interacciones de alto orden. ¿Cuáles efectos parecen estar activos?
b) Determine el mejor análisis de varianza e interprételo.
c) Obtenga las gráficas de los efectos que resultaron importantes en el ANOVA e interprételas.
d) Determine el mejor tratamiento.
e) Interprete con detalle las interacciones *AD* y *AE*, considerando que el factor *D* es el resultado de un proceso anterior y que actualmente no es posible fijarlo en un nivel, y que el factor *E* es una característica de los materiales de un proveedor externo; entonces, ¿cuáles son sus recomendaciones para operar el proceso y para acciones de mejora a futuro?
f) Verifique los supuestos del modelo. ¿Qué puede concluir del análisis de residuos?
g) ¿La forma especial de la gráfica de residuos contra predichos, afecta las conclusiones a la que llegó antes?
h) ¿Es pertinente colapsar este diseño en un factorial 2^4 con dos réplicas? Si la respuesta es positiva, hágalo.
i) ¿Se puede colapsar en un 2^3 con cuatro réplicas?
j) Antes del experimento se creía firmemente que el sistema de luz tradicional (fibra) ya era obsoleto, y que era parte de las causas principales del problema, por lo tanto habría que invertir en mejor tecnología (incandescente). Dados los resultados del experimento, ¿qué puede decir sobre esta "firme" creencia?
k) ¿Qué puede comentar sobre el hecho de que en más de la mitad de los tratamientos se haya tenido cero defectos?

15. En una planta de fuerza se corrió un experimento factorial 2^4 con repeticiones al centro, con el objetivo de hacer más eficaz la operación de la máquina de absorción. La eficacia de la máquina se mide en toneladas de refrigeración entre el flujo de vapor (*tr*/*fv*). Los factores a controlar fueron: flujo de vapor (*A*), temperatura de agua helada (*B*), temperatura de agua de enfriamiento (*C*), presión diferencial (*D*). El diseño, escrito en orden aleatorio y en las unidades originales, es el que se muestra más adelante.
a) Determine el mejor ANOVA para estos datos.
b) ¿Cuál porcentaje de la variación observada es explicado por el mejor ANOVA?
c) Realice la prueba de falta de ajuste. ¿Qué nos indica la presencia de efectos de curvatura?
d) Grafique los efectos significativos.
e) Determine el mejor tratamiento y haga la predicción de la eficacia esperada sobre él.

f) Verifique los supuestos de normalidad, varianza constante e independencia.

A (t/h)	*B* (°C)	*C* (°C)	*D* (kg/cm²)	Eficacia (tr/fv)
4.5	6.5	23	1.1	99
3.25	5.25	25	1.4	105
2.0	4.0	23	1.7	99
2.0	4.0	27	1.7	79
4.5	6.5	27	1.7	86
2.0	6.5	27	1.1	85
4.5	4.0	23	1.1	90
2.0	4.0	23	1.1	95
4.5	4.0	27	1.7	79
4.5	4.0	27	1.1	82
4.5	6.5	27	1.1	83
4.5	6.5	23	1.7	97
3.25	5.25	25	1.4	101
3.25	5.25	25	1.4	98
2.0	6.5	23	1.7	108
2.0	6.5	23	1.1	111
2.0	4.0	27	1.1	89
4.5	4.0	23	1.7	91
2.0	6.5	27	1.7	88

16. Una de las preocupaciones permanentes en la industria tequilera es obtener altos niveles de rendimiento, el cual puede depender de factores de control como: presión (*A*), tiempo de cocimiento (*B*) y tiempo de reposo (*C*). Se decide realizar un experimento factorial a nivel piloto para investigar si estos factores tienen efecto sobre el rendimiento, medido éste por la eficiencia en mieles (EM) y la eficiencia en agave cocido (EAC). Los resultados obtenidos en una sola réplica del diseño se muestran a continuación:

A	*B*	*C*	EM	EAC
−1	−1	−1	78.8	93.90
1	−1	−1	94.0	97.30
−1	1	−1	93.2	95.50
1	1	−1	95.9	96.91
−1	−1	1	90.6	94.60
1	−1	1	97.9	98.60
−1	1	1	95.7	94.00
1	1	1	90.5	95.80

a) Haga una inspección visual de los datos y con base en ella establezca una conjetura sobre si los factores afectan o no a las variables de respuesta estudiadas.
b) ¿Qué diseño es el que se ha corrido?
c) Analice EM con el gráfico de Daniel (gráfica de efectos en papel normal) y el diagrama de Pareto.
d) A partir de lo anterior construya un error, obtenga el ANOVA para tal variable de respuesta. ¿Cuáles efectos están activos?
e) ¿Cómo afecta en el análisis el hecho de no haber replicado el experimento?
f) Determine el mejor tratamiento, si es que lo hay, para EM.

g) Repita los incisos anteriores para la otra variable de respuesta, para cada una de las variables de rendimiento.

h) ¿Cuál es una posible causa de la poca variación observada en los datos de la variable EAC?

17. Una de las fallas más importantes en la línea de empaque de un producto es la calidad de las etiquetas. Un equipo de mejora decide atacar este problema mediante diseño de experimentos. Para ello eligen una de las impresoras a la cual se le pueden manipular los factores: velocidad, temperatura, tensión y tipo de etiqueta. Los niveles utilizados con cada factor fueron:

Factor	Nivel bajo	Centro	Nivel alto
Velocidad	baja	media	alta
Temperatura	5	13	21
Tensión	4	8	12
Tipo de etiqueta	esmaltada	otra	mate

El diseño factorial utilizado fue un 2^4 con repeticiones al centro. En cada combinación del experimento se imprimieron 20 etiquetas y se contabiliza como variable de respuesta en número de impresiones rechazadas. Los resultados observados, listados en orden aleatorio, fueron los que se muestran en la siguiente tabla.

¿Código?	Temperatura	Velocidad	Etiqueta	Tensión	No pasan
	1	1	1	1	20
	–1	1	–1	1	20
	1	1	1	–1	19
	1	–1	1	–1	9
	0	0	0	0	20
	1	1	–1	–1	3
	1	1	–1	1	20
	–1	–1	1	–1	20
	–1	1	1	1	20
	–1	–1	–1	1	20
	–1	–1	–1	–1	20
	1	–1	1	1	7
	–1	–1	1	1	20
	–1	1	–1	–1	20
	1	–1	–1	1	0
	1	–1	–1	–1	5
	–1	1	1	–1	20
	0	0	0	0	20

a) Utilice la notación de Yates y anote en la primera columna de la tabla el código correspondiente a cada una de las corridas, y asegúrese de que se corrieron todos los tratamientos correspondientes al diseño empleado.

b) Encuentre el mejor ANOVA para estos datos.
c) Grafique los efectos significativos e interprételos para determinar el tratamiento ganador.
d) Estime el número de etiquetas que se espera sigan sin pasar en el mejor tratamiento.
e) Verifique supuestos. ¿Hay algún problema potencial?

18. Se hace un experimento para mejorar el *rendimiento* de un proceso, controlando cuatro factores en dos niveles cada uno. Se corre una réplica de un diseño factorial 2^4, con los factores tiempo (*A*), concentración (*B*), presión (*C*) y temperatura (*D*), y los resultados son los siguientes:

	A_0				A_1			
	B_0		B_1		B_0		B_1	
	C_0	C_1	C_0	C_1	C_0	C_1	C_0	C_1
D_0	12	17	13	20	18	15	16	15
D_1	10	19	13	17	25	21	24	23

a) Analice estos datos con el uso de todos los criterios existentes para encontrar el mejor ANOVA. En las figuras considere de entrada los 15 efectos posibles.
b) ¿Cuáles efectos están activos?
c) Determine el mejor tratamiento.
d) Prediga el rendimiento esperado en el mejor tratamiento y dé un intervalo de confianza para el rendimiento futuro.
e) Compruebe los supuestos del modelo.
f) ¿Puede este diseño colapsarse en uno 2^3 con dos réplicas? De ser posible, hágalo y repita los incisos anteriores para este nuevo diseño.

19. Se realiza un experimento para mejorar el rendimiento de un proceso químico. Se seleccionan cuatro factores y se hace sólo una repetición en cada tratamiento, con los siguientes resultados:

Trat	Rend	Trat	Rend
(1)	90	*d*	98
a	74	*ad*	72
b	81	*bd*	87
ab	83	*abd*	85
c	77	*cd*	99
ac	81	*acd*	79
bc	88	*bcd*	87
abc	73	*abcd*	80

a) Incluyendo las interacciones de mayor orden, analice los efectos con gráfica de Pareto y con la gráfica de papel normal para efectos. ¿Qué efectos parecen significativos?
b) A partir de lo anterior construya un error y haga un análisis de varianza. ¿Qué efectos son significativos?
c) Si hay una interacción importante, interprétela con detalle.
d) ¿Cuál es el mejor tratamiento?

e) Verifique los supuestos del modelo.
f) Si algún factor no tiene ningún tipo de efecto, colapse el diseño en un $2^3(n = 2)$. Repita los incisos previos y compare los resultados.

20. En el ejemplo de la sección "Cuando casi ningún efecto es significativo: un ejemplo" de este capítulo, uno de los análisis alternativos fue colapsar un diseño 2^5 en uno 2^3 con cuatro réplicas. Para este diseño:
a) Haga el análisis completo del diseño colapsado, interprete los efectos activos y verifique supuestos.
b) ¿Obtuvo las mismas conclusiones que se consiguieron con el análisis del diseño sin colapsar? ¿A qué lo atribuye?

21. En una empresa del área electrónica se quieren minimizar los problemas generados en el proceso conocido como "Soldadora de ola". Los defectos que se quieren reducir son insuficiencias de soldadura en las tarjetas. Los factores y niveles que inicialmente se decide estudiar son: velocidad de *conveyor* (4 y 7 pies/minuto), temperatura de precalentado (80 y 120°C), y temperatura de soldadura (470 y 500°C). Debido a que el proceso es muy rápido (se suelda una tarjeta cada 10 a 15 segundos) se decide soldar en cada condición de prueba 25 tarjetas. La variable de respuesta es la cantidad de insuficiencias detectadas en los diferentes puntos de soldadura de las 25 tarjetas. Se hicieron dos réplicas. La matriz de diseño y los datos obtenidos se muestran a continuación:

Velocidad	Precalentado	Soldadura	Insuficiencias	
4	80	470	29	25
7	80	470	110	110
4	120	470	23	27
7	120	470	77	59
4	80	500	12	44
7	80	500	146	162
4	120	500	51	35
7	120	500	42	48

a) Haga un análisis completo y determine los efectos más importantes, el ANOVA y el análisis de residuos.
b) Al parecer, la interacción velocidad-precalentado es importante, de ser así realice una interpretación detallada de tal interacción en términos físicos.
c) ¿Cuáles serían las condiciones de operación del proceso que podrían utilizarse para reducir la cantidad de insuficiencias? Analice las opciones disponibles.

22. El tequila es una bebida que está sujeta a una norma oficial mexicana, y conforme a ésta se debe cumplir con ciertas especificaciones físico-químicas. En un laboratorio de investigación, mediante un diseño factorial 2^5 no replicado, se estudió la influencia de diversos factores sobre la producción de alcoholes superiores en la etapa de fermentación (Pinal *et al*., 1997). Los factores estudiados y los niveles fueron: tipo de cepa, *A*(1, 2), temperatura, *B*(30, 35°C), fuente de nitrógeno, *C*$(NH_4)_2SO_4$ y urea-, relación carbono/nitrógeno, *D*(62/1, 188/1) y porcentaje de inóculo, *E*(5 y 10%). En la siguiente tabla se muestran los resultados obtenidos en cuanto a alcohol isoamílico (mg/L), que es parte de los alcoholes superiores.

(1) = 21.4	d = 42.5	e = 32.9	de = 54.0
a = 16.8	ad = 21.0	ae = 17.5	ade = 21.8
b = 29.3	bd = 79.1	be = 30.0	bde = 79.9
ab = 12.7	abd = 20.0	abe = 24.1	$abde$ = 31.5
c = 27.5	cd = 48.6	ce = 26.7	cde = 47.9
ac = 22.9	acd = 27.1	ace = 11.4	$acde$ = 15.6
bc = 35.4	bcd = 85.2	bce = 23.9	$bcde$ = 73.8
abc = 18.8	$abcd$ = 26.1	$abce$ = 18.0	$abcde$ = 25.4

a) Dibuje el diagrama de Pareto y el gráfico de Daniel considerando todas las interacciones de alto orden. ¿Cuáles efectos parecen estar activos?
b) Determine el mejor análisis de varianza e interprételo.
c) Obtenga las gráficas de los efectos que resultaron importantes en el ANOVA, e interprételas con detalle.
d) Determine los tratamientos que minimizan y maximizan la variable de respuesta.
e) Verifique los supuestos del modelo. ¿Qué puede concluir del análisis de residuos?
f) ¿Es pertinente colapsar este diseño en un factorial 2^4 con dos réplicas? Si la respuesta es positiva, hágalo.

23. Interesa estudiar el efecto de la temperatura y del tiempo de remojo sobre la corrosión en navajas de rasurar. Cada observación se obtiene de someter una navaja al tratamiento y posteriormente ponerla en una cámara con 98% de humedad por 48 horas; la respuesta medida es el porcentaje de superficie que exhibe corrosión. Los resultados obtenidos para ocho cuchillas se muestran en la siguiente tabla:

	Tiempo (min.)	
Temperatura (°F)	5 min.	60 min.
650	30%, 20%	25%, 30%
690	75%, 85%	95%, 90%

a) ¿El tiempo de remojo y la temperatura afectan la corrosión de las navajas?
b) Dibuje las gráficas de los efectos activos y obtenga el tratamiento ganador.

24. Haga el análisis de los datos del ejemplo 6.1 (corriente de fuga) y obtenga la tabla de ANOVA 6.22. Compruebe los cálculos de las sumas de cuadrados. Colapse el diseño en el sentido del factor *B* y escriba el diseño resultante. Analícelo y saque conclusiones.

25. En una fábrica de dulces hay problemas por la alta variabilidad de su peso. El dulce se forma vertiendo en moldes con varias cavidades, mediante un proceso de dosificado continuo. Los factores que se desea estudiar son: viscosidad de dulce líquido (*A*: −1, 1), velocidad del rotor 1(*B*: −1, 1), velocidad rotor 2(*C*: −1, 1). Se decide correr un diseño 2^3 con dos réplicas y un punto al centro por réplica. Las corridas experimentales se hicieron a nivel proceso; cada una consistió en poner a trabajar el dosificador y después de un tiempo se tomó una muestra de 15 dulces, que fueron pesados de manera individual. Con esos 15 datos se obtuvo la media (*Y*1) y la desviación estándar (*Y*2) para cada corrida. Los datos obtenidos se muestran a continuación:

A	B	C	Y_1	Y_2
–1	–1	–1	5.747	0.05
–1	–1	–1	5.6875	0.054
1	–1	–1	5.697	0.05
1	–1	–1	5.694	0.05
–1	1	–1	6.3905	0.07
–1	1	–1	6.347	0.05
1	1	–1	6.3005	0.072
1	1	–1	6.45	0.045
–1	–1	1	6.2065	0.054
–1	–1	1	6.425	0.073
1	–1	1	5.691	0.054
1	–1	1	5.625	0.051
–1	1	1	6.394	0.052
–1	1	1	6.4095	0.062
1	1	1	5.6016	0.053
1	1	1	5.6565	0.048

a) Haga un análisis de varianza para cada una de las variables de respuesta y destaque los aspectos más relevantes.
b) Realice un análisis detallado de los residuales para ambas variables.
c) ¿Los factores controlados tienen influencia significativa en la variabilidad del peso? Argumente su respuesta.
d) ¿Si se quiere que el peso sea de seis, hay algún tratamiento que lo garantice?
e) Con el modelo de regresión ajustado estime ¿cuál sería el valor de *Y*1 en el centro de la región experimental?
f) Con la respuesta del punto anterior, ¿hubiese sido interesante correr punto al centro en este diseño? Argumente su respuesta.

Capítulo 7

Diseños factoriales 3^k y factoriales mixtos

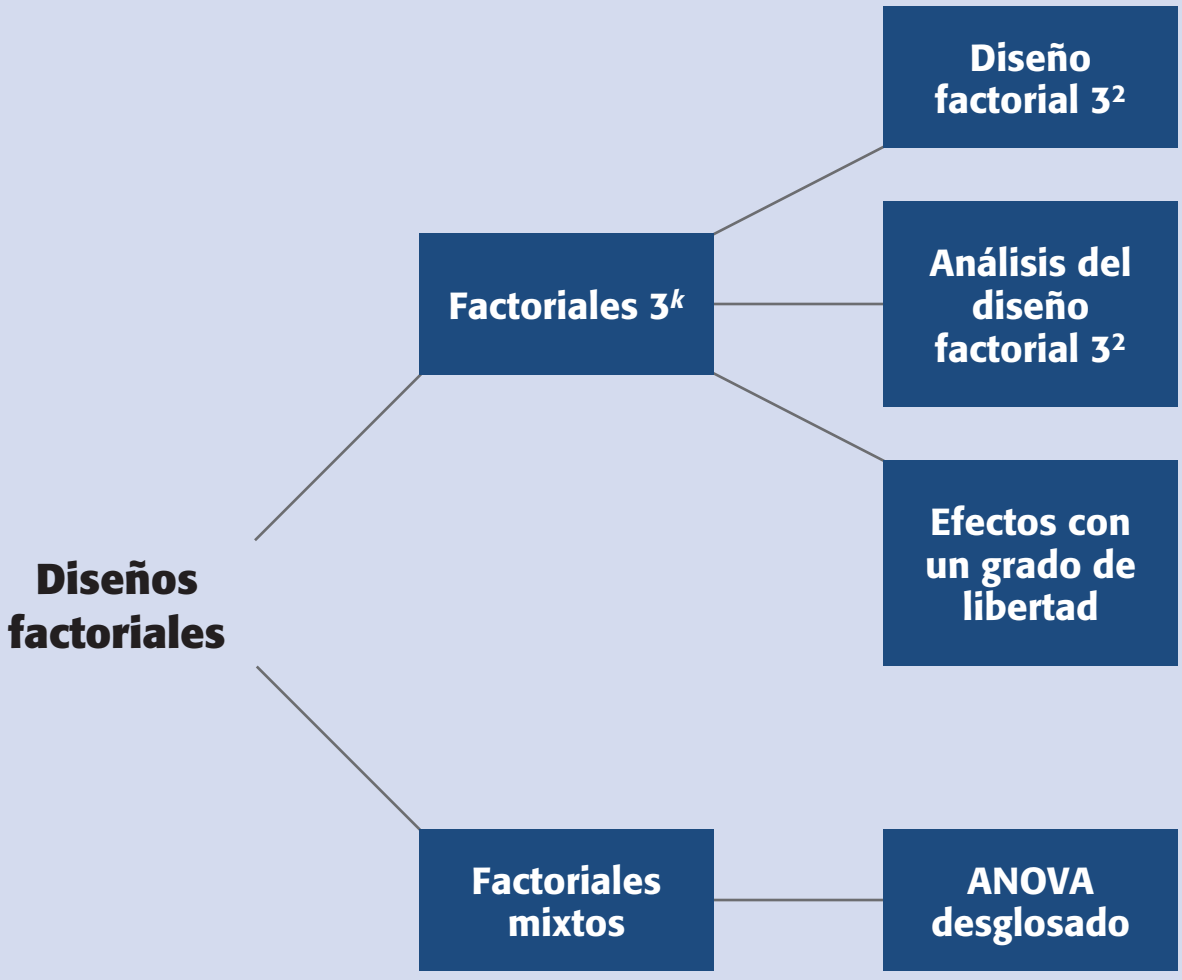
Sumario

- Diseños factoriales 3^k
- Factores mixtos
- Uso de software estadístico

Objetivos de aprendizaje

- Diferenciar entre diseños factoriales 2^k y 3^k para determinar en qué situación es apropiado cada uno.
- Identificar el diseño factorial mixto y sus características.
- Analizar el diseño factorial mixto y el 3^k, desglosando su ANOVA hasta efectos con un grado de libertad.

Mapa conceptual



Conceptos clave

- ANOVA desglosado
- Curvatura pura
- Diseño factorial 3^k
- Factorial mixto
- Parte lineal de un efecto

En el capítulo anterior vimos con detalle un caso importante de los diseños factoriales que permiten estudiar los diseños 2^k que es el efecto simultáneo de varios factores sobre una o más variables de respuesta. En este capítulo estudiaremos otros diseños factoriales, en los que al menos uno de los factores tiene más de dos niveles. Estos diseños, aunque menos utilizados que los 2^k, también son útiles en muchas situaciones prácticas. Para iniciar recordemos que en un diseño factorial general la matriz de diseño que contiene el total de tratamientos a evaluar se obtiene encontrando todas las posibles combinaciones que se pueden formar con los niveles de los k factores de interés, donde cada factor tiene al menos dos niveles. Por ejemplo, el diseño factorial $4 \times 4 \times 3$ estudia tres factores: dos con cuatro niveles y el tercer factor con tres niveles, esto hace que la matriz de diseño esté formada por un total de $4 \times 4 \times 3 = 48$ tratamientos. Iniciamos este capítulo con el caso especial donde cada factor estudiado tiene tres niveles, antes de discutir el caso donde los factores usan diferentes cantidades de niveles (factoriales mixtos).

Diseños factoriales 3^k

Diseño factorial 3^k
Modelo que considera k factores con tres niveles cada uno y tiene 3^k tratamientos. Este diseño requiere de mayor cantidad de pruebas que el diseño 2^k.

El *diseño factorial* 3^k considera k factores con tres niveles cada uno y tiene 3^k tratamientos. La primera desventaja de los diseños 3^k es que al aplicarse requieren mayor cantidad de pruebas que el diseño 2^k. Por ejemplo, si se quieren estudiar cuatro factores, y se considera sólo una repetición, el diseño 3^4 requiere en total 81 pruebas, una para cada tratamiento; mientras que el diseño 2^4 sólo necesita 16 pruebas. De esta forma, cuando se tienen muchos factores, cuatro o más, prácticamente es prohibitivo pensar en correr un diseño 3^k. Sin embargo, cuando se tienen pocos factores, tres como máximo, o en el peor de los casos cuatro, hay algunas situaciones prácticas en las que el diseño 3^k es una buena alternativa. De manera específica, hay dos razones que hacen viable el diseño 3^k:

- Se tienen factores de tipo continuo e interesa estudiar efectos cuadráticos como A^2, B^2,…, A^2B, B^2A, A^2B^2,… (efectos de curvatura). Esto se hace cuando se cree que la variable de respuesta no es lineal, ni aproximadamente lineal en el rango de variación de los factores estudiados (ver figura 7.1*b*).
- Los factores son categóricos o discretos y de manera natural tienen tres niveles cada uno. Esto ocurre en factores como tipo de material, diferentes medios de cultivos, factor máquina, etc., los cuales pueden tener naturalmente tres niveles. Por ejemplo, si uno de los factores es un *reactivo* del cual existen tres marcas, resulta natural que el experimentador quiera probar las tres marcas con la idea de comparar su desempeño en las diferentes combinaciones de los factores restantes.

Ahora, profundizando en la posible existencia de curvatura, consideremos un factor X de tipo continuo que se supone tiene efecto sobre la respuesta Y. Dicho efecto desconocido podría ser de tipo lineal, al menos en forma aproximada, como en la figura *7.1a*, o de tipo cuadrático como en la figura *7.1b*. En esta misma figura se observa que para estudiar un efecto lineal, o aproximadamente lineal, basta con probar el factor X en dos niveles, mientras que para estudiar un efecto cuadrático son

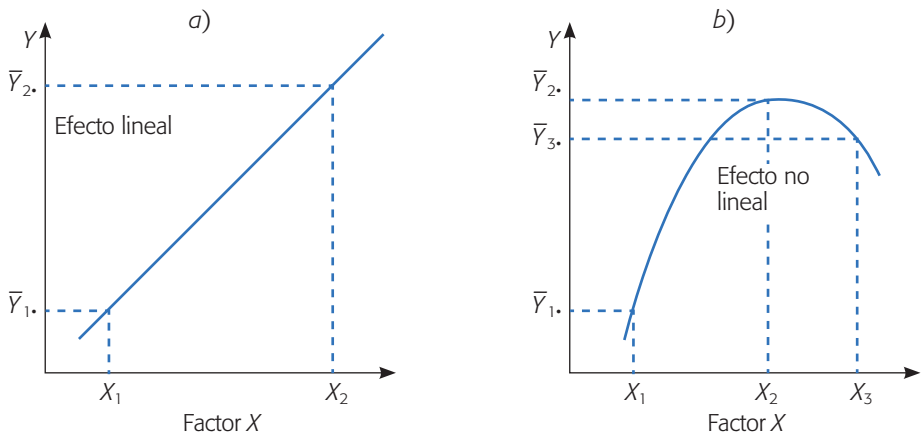


Figura 7.1 Ejemplos de efecto lineal y efecto con curvatura.

necesarios al menos tres niveles del factor *X*. El problema es que de antemano no se sabe cómo será el efecto del factor y a veces se corre el riesgo de utilizar dos niveles, cuando en realidad se requerían de al menos tres. Entonces, se recomienda contar con una estrategia experimental que permita detectar de manera económica la presencia de curvatura. Ése es precisamente el propósito de agregar puntos al centro en un diseño 2^k. Sin embargo, con esta estrategia sólo se detecta si hay curvatura, pero no es posible estudiarla (modelarla).

El diseño factorial 3^k es una de las alternativas experimentales que permite estudiar efectos de curvatura, además de efectos lineales y de interacción. Otros diseños, que de hecho son más utilizados y recomendados para ese fin, son el *diseño de Box-Benhken* y el *diseño central compuesto* que se presentan en el capítulo 12.

Diseño factorial 3^2

Este diseño consiste en $3^2 = 9$ tratamientos diferentes, que corresponden a todas las posibles maneras en que se pueden combinar dos factores en tres niveles cada uno. Sean *A* y *B* los factores, cada uno con tres niveles, a los cuales se les suele llamar *bajo*, *medio* y *alto*. Los nueve tratamientos se pueden escribir de varias maneras, algunas de las cuales se muestran en la tabla 7.1.

Tabla 7.1 Diseño factorial 3^2 en tres notaciones útiles.

Tratamiento	*A*	*B*	*A*	*B*	*A*	*B*
1	bajo	bajo	–1	–1	0	0
2	medio	bajo	0	–1	1	0
3	alto	bajo	1	–1	2	0
4	bajo	medio	–1	0	0	1
5	medio	medio	0	0	1	1
6	alto	medio	1	0	2	1
7	bajo	alto	–1	1	0	2
8	medio	alto	0	1	1	2
9	alto	alto	1	1	2	2

Los nueve puntos de diseño se pueden representar en forma geométrica sobre el cuadrado de la figura 7.2, que delimita la región experimental.

Análisis del diseño factorial 3^2

El modelo estadístico para el diseño 3^2 se puede escribir considerando el efecto individual de cada factor y de la interacción entre ambos, y queda como sigue:

$$Y_{ijk} = \mu + \gamma_i + \delta_j + (\gamma\delta)_{ij} + \varepsilon_{ijk} \qquad \textbf{(7.1)}$$

$$\text{con } i = 1, 2, 3; j = 1, 2, 3; k = 1, \ldots, n$$

donde γ_i es el efecto del factor A en su nivel i, δ_j representa el efecto del factor B en su nivel j $(y\delta)_{ij}$ es el efecto de interacción de ambos en los niveles ij y n es el número de repeticiones de cada tratamiento. En consecuencia, las hipótesis que se desean probar son: $H_0 : \gamma_i = 0$ (no hay efecto significativo del factor A sobre la variable de respuesta), $H_0 : \delta = 0$ (no hay efecto del factor B sobre la variable de respuesta) y $H_0 : (\gamma\delta)_{ij} = 0$ (no hay efecto de interacción de los factores A y B sobre la variable de respuesta). Estas hipótesis se probarán con el ANOVA; para ello, las sumas de cuadrados para los tres efectos incluidos en la ecuación (7.1) están dadas por:

$$SC_A = \sum_{i=1}^{3} \frac{Y_{i\bullet\bullet}^2}{3n} - \frac{Y_{\bullet\bullet\bullet}^2}{n3^2}$$

$$SC_B = \sum_{j=1}^{3} \frac{Y_{\bullet j\bullet}^2}{3n} - \frac{Y_{\bullet\bullet\bullet}^2}{n3^2}$$

$$SC_{AB} = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{Y_{ij\bullet}^2}{n} - \frac{Y_{\bullet\bullet\bullet}^2}{n3^2} - SC_A - SC_B$$

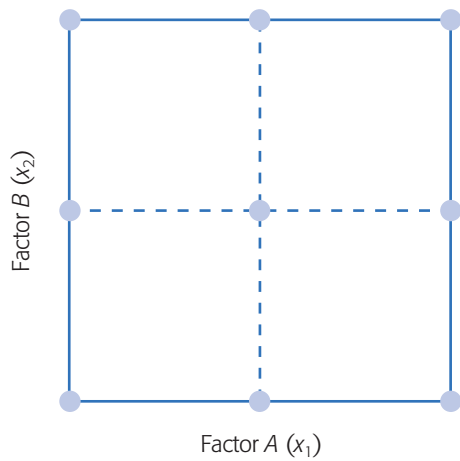


Figura 7.2 Representación en el plano del diseño 3^2.

La suma de cuadrados total se obtiene con:

$$SC_T = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{n} Y_{ijk}^2 - \frac{Y_{...}^2}{n3^2} \qquad (7.2)$$

y el error aleatorio se calcula con la diferencia,

$$SC_E = SC_T - SC_{AB} - SC_A - SC_B$$

Los grados de libertad asociados con cada suma de cuadrados de esta última relación son, respectivamente:

$$3^2(n-1) = (n3^2 - 1) - (3-1)\,(3-1) - (3-1) - (3-1)$$

El bosquejo del análisis de varianza para el diseño 3^2 se muestra en la tabla 7.2. Observe que este diseño también requiere al menos de dos repeticiones para que haya grados de libertad para el error. Note que las sumas de cuadrados de los efectos A y B tienen 2 grados de libertad (número de niveles menos 1) y que los grados de libertad para la interacción AB se obtienen con el producto $2 \times 2 = 4$. Si el valor-p para un efecto es menor que 0.05, entonces se rechaza la correspondiente hipótesis nula, y se concluye que tal fuente de variación afecta de manera significativa a la variable de respuesta.

Descomposición a efectos con 1 grado de libertad. El ANOVA de la tabla 7.2 considera los efectos A, B y AB de manera global, es decir, sin especificar si influyen de manera lineal, cuadrática o de ambas formas. Las sumas de cuadrados de cada efecto se pueden descomponer en sumas de cuadrados con un grado de libertad. Por ejemplo, la suma de cuadrados del efecto A con 2 grados de libertad se puede desglosar en los componentes A_L y A^2 (efecto lineal y cuadrático) cada uno con un grado de libertad. El subíndice L indica que es la parte lineal del efecto global A, y el exponente denota la parte cuadrática. Asimismo, la suma de cuadrados de AB dada en la tabla 7.2 se puede partir en cuatro componentes con un grado de libertad: A_LB_L, A_LB^2, A^2B_L y A^2B_L. Así, para tener información más detallada de cómo es que afectan los factores, es necesario separar los ocho componentes con un grado de libertad y construir un ANOVA que muestre la significancia de cada uno de ellos. Este ANOVA desglosado se muestra en la tabla 7.3.

Tabla 7.2 ANOVA para el diseño 3^2.

FV	SC	GL	CM	F_0	Valor-p
A	SC_A	2	CM_A	CM_A/CM_E	$P(F > F_0)$
B	SC_B	2	CM_B	CM_B/CM_E	$P(F > F_0)$
AB	SC_{AB}	4	CM_{AB}	CM_{AB}/CM_E	$P(F > F_0)$
Error	SC_E	$3^2\,(n-1)$	CM_E		
Total	SC_T	$n3^k - 1$			

Tabla 7.3 ANOVA desglosado para el diseño 3^2.

FV	SC	GL	CM	F_0	Valor-*p*
A_L	SC_{A_L}	1	CM_{A_L}	CM_{A_L}/CM_E	$P(F > F_0)$
B_L	SC_{B_L}	1	CM_{B_L}	CM_{B_L}/CM_E	$P(F > F_0)$
A^2	SC_{A^2}	1	CM_{A^2}	CM_{A^2}/CM_E	$P(F > F_0)$
B^2	SC_{B^2}	1	CM_{B^2}	CM_{B^2}/CM_E	$P(F > F_0)$
$A_L B_L$	$SC_{A_L B_L}$	1	$CM_{A_L B_L}$	$CM_{A_L B_L}/CM_E$	$P(F > F_0)$
$A^2 B_L$	$SC_{A^2 B_L}$	1	$CM_{A^2 B_L}$	$CM_{A^2 B_L}/CM_E$	$P(F > F_0)$
$A_L B^2$	$SC_{A_L B^2}$	1	$CM_{A_L B^2}$	$CM_{A_L B^2}/CM_E$	$P(F > F_0)$
$A^2 B^2$	$SC_{A^2 B^2}$	1	$CM_{A^2 B^2}$	$CM_{A^2 B^2}/CM_E$	$P(F > F_0)$
Error	SC_E	$3^2 (n-1)$	CM_E		
Total	SC_T	$n3^2 - 1$			

Debemos señalar que en muchos casos el riesgo de tomar una decisión incorrecta con base en la tabla 7.2, aun sin desglosar a los componentes, es pequeño o nulo, ya que al graficar los efectos globales A, B y AB considerando los tres niveles de cada factor, se clarifica lo que ocurre con ellos (véase figura 7.3), y se toma en cuenta al momento de interpretarlos. Es decir, el desglose dado en la tabla 7.3 puede omitirse en la mayoría de las situaciones prácticas, pero aquí lo hemos hecho para que el lector comprenda mejor las bondades del diseño 3^2. En la figura 7.3 se muestra un ejemplo de efecto de A, la línea une las medias observadas $(\bar{Y}_{1\cdot\cdot}, \bar{Y}_{2\cdot\cdot}, \bar{Y}_{3\cdot\cdot})$ en cada nivel del factor A, respectivamente.

En la práctica, un efecto puede ser lineal, como en la figura 7.3*a*, ya que su componente más activo es la parte lineal A_L, o también puede ser prácticamente cuadrático, como en la figura 7.3*b*, ya que su componente más activo es la parte cuadrática A^2. Como se observa, en cualquier caso la representación del efecto permite comprender cómo está actuando físicamente el factor sobre la variable de respuesta. Además, a partir de las gráficas es muy sencillo, en ambos casos, localizar el mejor tratamiento.

Cuando se tiene un efecto que parece ser cuadrático A^2, no es suficiente ver la gráfica correspondiente para saber si el efecto cuadrático es significativo, adicionalmente se requiere respaldarlo con el análisis de varianza. Para ello, se requiere estimar la suma de cuadrados SC_{A^2}, como se muestra en el ANOVA desglosado de la tabla 7.3. Sean $Y_{1\cdot\cdot}$, $Y_{2\cdot\cdot}$ y $Y_{3\cdot\cdot}$ los totales observados de la variable de respuesta en los niveles bajo, medio y alto del factor A, respectivamente. Recordemos que en los capítulos 3 y 6 se planteó que una combinación lineal de la forma $C_1Y_{1\cdot\cdot} + C_2Y_{2\cdot\cdot} + C_3Y_{3\cdot\cdot}$ con $C_1 + C_2 + C_3 = 0$ se llama contraste, a partir de esto veremos que los coeficientes del contraste A^2 son (1, –2, 1), para medir con este contraste la diferencia entre la respuesta observada en el nivel medio y la respuesta en los niveles bajo y alto.

Curvatura pura
Se obtiene al graficar el efecto de un factor con tres niveles, y calcular la diferencia entre la pendiente del segundo y la del primer segmento de la gráfica.

Para deducir los coeficientes (1, –2, 1), notemos que en la figura 7.3*a* la *curvatura pura* se puede definir como la diferencia entre la pendiente del segundo y la del primer segmento. De esto se deduce que a mayor diferencia en tales pendientes, mayor es la curvatura (ver por ejemplo el caso de una parábola en la figura 7.3*b*), y si no hay diferencia en las pendientes no existe curvatura. Recordemos que, en general,

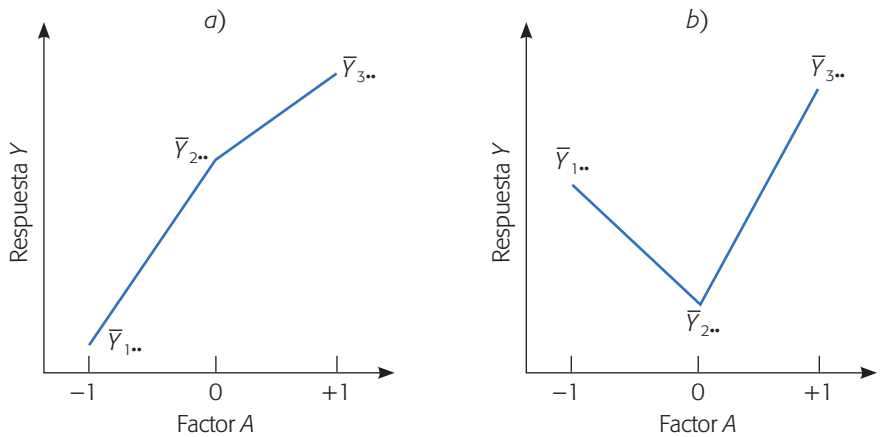


Figura 7.3 Gráfica de efectos predominando: *a*) parte lineal A_L, y *b*) parte cuadrática A^2.

para calcular pendientes de una línea recta se divide el incremento en el eje vertical contra el incremento en el eje horizontal. Tomando en cuenta esto, la pendiente del segundo segmento es $(Y_{3..} - Y_{2..})/1$, ya que el incremento en el eje horizontal es uno, y la pendiente del primer segmento es $(Y_{2..} - Y_{1..})/1$. Así, las diferencias de estas pendientes o contraste de la curvatura pura A^2 queda como:

$$\begin{aligned} \text{Contraste } A^2 &= (Y_{3..} - Y_{2..}) - (Y_{2..} - Y_{1..}) \\ &= (1)Y_{3..} + (-2)Y_{2..} + (1)Y_{1..} \end{aligned} \tag{7.3}$$

Parte lineal de un efecto
Mide el cambio en la respuesta obtenida en los niveles extremos de un factor.

que tiene los coeficientes (1, –2, 1). De manera análoga, la *parte lineal del efecto A* se puede explicar como la suma de las pendientes de los segmentos, de aquí que el contraste del efecto lineal está dado por:

$$\begin{aligned} \text{Contraste } A_L &= (Y_{3..} - Y_{2..}) + (Y_{2..} - Y_{1..}) \\ &= (1)Y_{3..} + (0)Y_{2..} + (-1)Y_{1..} \end{aligned} \tag{7.4}$$

Note que al sumar las pendientes se ignora la curvatura y queda como resultado el cambio observado entre los niveles –1 y 1; lo que pasó en el nivel intermedio no importa para fines del efecto lineal de *A*.

Del contraste para la curvatura dado en la ecuación 7.3 se obtiene su correspondiente suma de cuadrados con la fórmula:

$$SC_{A^2} = \frac{(\text{Contraste } A^2)^2}{n\sum_{i=1}^{3}\sum_{j=1}^{3} c_{ij}^2} \tag{7.5}$$

donde n es el número de réplicas, $C_{1j} = 1$, $C_{2j} = -2$ y $C_{3j} = 1$, y tiene un grado de libertad.

El efecto lineal de *A* se obtiene utilizando su contraste con coeficientes (1, 0, –1) que compara el nivel alto con el nivel bajo. Para calcular la suma de cuadrados (SC_{A_L}) se emplean las mismas operaciones que hicimos para el efecto cuadrático puro, pero ahora usando el contraste dado por la relación de la ecuación 7.4.

Ejemplo 7.1

En un proceso de fabricación de cajas se utiliza pegamento; con la idea de mejorar el desempeño de las cajas se realiza un experimento para estudiar la fuerza de adhesión del pegamento en diferentes condiciones de humedad y temperatura. La variable de respuesta es la fuerza necesaria en libras para despegar la caja. Los datos obtenidos en cada una de las nueve combinaciones de un diseño factorial 3^2 con $n = 2$ réplicas, se muestran a continuación:

	***A*: Temperatura**			
***B*: Humedad**	**Frío**	**Ambiente**	**Caliente**	**Total**
50%	1.5, 1.2	3.5, 3.2	4.0, 4.2	17.6
70%	1.4, 1.3	2.9, 2.5	3.8, 3.4	15.3
90%	0.8, 1.2	1.8, 2.0	2.7, 3.0	11.5
Totales	7.4	15.9	21.1	44.4

En la tabla de datos se calculan de una vez los totales por renglón y por columnas, así como el total global, dado que a partir de ellos se pueden calcular los efectos y las sumas de cuadrados. Las sumas de cuadrados de los efectos están dadas por:

$$SC_A = \sum_{i=1}^{3} \frac{Y_{i\cdot\cdot}^2}{3n} - \frac{Y_{\cdots}^2}{n3^2} = \frac{(7.4^2 + 15.9^2 + 21.1^2)}{3\times 2} - \frac{44.4^2}{18} = 15.943$$

$$SC_B = \sum_{j=1}^{3} \frac{Y_{\cdot j\cdot}^2}{3n} - \frac{Y_{\cdots}^2}{n3^2} = \frac{(17.6^2 + 15.3^2 + 11.5^2)}{3\times 2} - \frac{44.4^2}{18} = 3.163 \qquad \textbf{(7.6)}$$

$$SC_{AB} = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{Y_{ij\cdot}^2}{n} - \frac{Y_{\cdots}^2}{n3^2} - SC_A - SC_B = \frac{(2.7^2 + 6.7^2 + \cdots + 5.7^2)}{2} - \frac{44.4^2}{18} - 15.943 - 3.163 = 0.694$$

donde las cantidades $Y_{ij\cdot}^2$ son los cuadrados del total en cada tratamiento (cada celda de la tabla de resultados). La suma de cuadrados total resulta ser,

$$SC_T = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{n} Y_{ijk}^2 - \frac{Y_{\cdots}^2}{n3^2} = (1.5^2 + 1.2^2 + \ldots + 2.7^2 + 3.0^2) - \frac{44.4^2}{18} = 20.22$$

y finalmente la suma de cuadrados del error es:

$$SC_E = SC_T - SC_A - SC_B - SC_{AB} = 20.22 - 15.943 - 3.163 - 0.694 = 0.42$$

Los grados de libertad de SC_A, SC_B y SC_{AB} son 2, 2 y 4, respectivamente. En total el experimento tiene $(2 \times 3^2) - 1 = 17$ grados de libertad, y entonces quedan $17 - 2 - 2 - 4 = 9$ grados de libertad para la SC_E. Con esta información se obtiene la tabla 7.4 de ANOVA, de la cual se concluye que sí hay efecto significativo de la temperatura (A), de la humedad (B) y de la interacción de ambas sobre Y. Además del F_0 se aprecia que A tiene un efecto mucho más importante, seguido por el efecto de B, y un efecto pequeño de AB. Pero de este ANOVA no es posible saber cuál de los componentes, cuadrático o lineal, con un grado de libertad de cada efecto es el que predomina.

Tabla 7.4 ANOVA sin desglosar, ejemplo 7.1.

FV	SC	GL	CM	F_0	Valor-*p*
A	15.94	2	7.97	169.57	0.0000
B	3.16	2	1.58	33.62	0.0000
AB	0.69	4	0.17	3.62	0.0347
Error	0.42	9	0.047		
Total	20.22	17			

En la figura 7.4 se grafican los efectos principales y el efecto de interacción. Se observa que en los efectos predominan sus componentes lineales, ya que las líneas tienen una curvatura apenas perceptible. Se observa que a mayor temperatura y menor humedad, más efectivo es el proceso de pegado. En la gráfica de interacción apenas se aprecia un pequeño efecto (las líneas tienen pendiente similar). De esta manera, para efectos prácticos, aquí terminaría el análisis; sin embargo, si el objetivo es verificar con más detalle los efectos cuadráticos y lineales, se desglosan los efectos con un grado de libertad. Cosa que haremos a continuación.

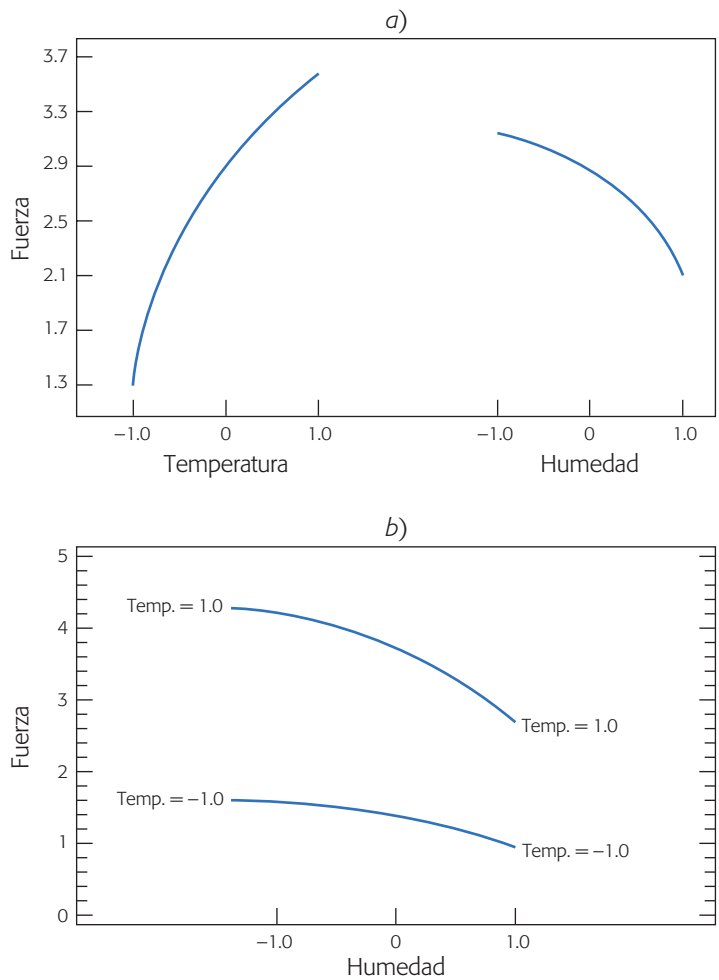


Figura 7.4 Efectos principales y efecto de interacción para el ejemplo de fabricación de cajas.

De acuerdo con las ecuaciones (7.3) a (7.5), se obtienen las siguientes sumas para el factor A:

$$SC_{A_L} = \frac{(-7.4+21.1)^2}{12} = 15.64$$
$$SC_{A^2} = \frac{(7.4-(2\times 15.9)+21.1)^2}{36} = 0.30 \qquad \textbf{(7.7)}$$

También para el factor B se obtiene que,

$$SC_{B_L} = \frac{(-17.6+11.5)^2}{12} = 3.10$$
$$SC_{B^2} = \frac{(17.6-(2\times 15.3)+11.5)^2}{36} = 0.062 \qquad \textbf{(7.8)}$$

Para descomponer la SC_{AB} se requiere el total dentro de cada combinación de niveles de los factores. En la tabla 7.5 se muestra el total en cada combinación, así como los coeficientes que corresponden a los contrastes. Con esta tabla se facilitan los cálculos de los componentes de la interacción. Observe que en la tabla también se incluyen los coeficientes para separar los efectos principales en sus dos componentes, mismos que se utilizaron para obtener las sumas dadas en las ecuaciones (7.7) y (7.8).

Al multiplicar la correspondiente columna de coeficientes de la tabla 7.5 por el total observado en cada combinación o tratamiento del diseño, se obtienen los componentes de la suma de cuadrados de la interacción.

$$SC_{A_LB_L} = \frac{(2.7-8.2-2.0+5.7)^2}{8} = 0.405 \qquad \textbf{(7.9)}$$
$$SC_{A_LB^2} = \frac{(-2.7+8.2+2\times 2.7-2\times 7.2-2.0+5.7)^2}{24} = 0.0017$$
$$SC_{A^2B_L} = \frac{(-2.7+2\times 6.7-8.2+2.0-2\times 3.8+5.7)^2}{24} = 0.282$$
$$SC_{A^2B^2} = \frac{(2.7-2\times 6.7+8.2-2\times 2.7+4\times 5.4-2\times 7.2+2.0-2\times 3.8+5.7)^2}{72} = 0.005$$

En la tabla 7.6 se muestra el ANOVA para los efectos desglosados con un grado de libertad. Es fácil comprobar que las sumas de cuadrados con un grado de libertad constituyen la suma de cuadrados de los efectos originales sin descomponer, es decir, se verifica que $SC_{A_L} + SC_{A^2} = SC_A$, $SC_{B_L} + SC_{B^2} = SC_B$ y $SC_{A_LB_L} + SC_{A_LB^2} + SC_{A^2B_L} + SC_{A^2B^2} = SC_{AB}$. Al comparar la tabla 7.4 del ANOVA sin desglosar con la tabla 7.6 del ANOVA desglosado, se concluye que básicamente están activas las partes lineales de los efectos principales y un ligero efecto de interacción y de la componente cuadrática de A^2. Como se vio desde las gráficas de efectos dadas en la figura 7.4.

Tabla 7.5 Coeficientes para calcular los contrastes en el factorial 3^2, ejemplo 7.1.

A	B	Total	A_L	A^2	B_L	B^2	A_LB_L	A_LB^2	A^2B_L	A^2B^2
–1	–1	2.7	–1	1	–1	1	1	–1	–1	1
0	–1	6.7	0	–2	–1	1	0	0	2	–2
1	–1	8.2	1	1	–1	1	–1	1	–1	1
–1	0	2.7	–1	1	0	–2	0	2	0	–2
0	0	5.4	0	–2	0	–2	0	0	0	4
1	0	7.2	1	1	0	–2	0	–2	0	–2
–1	1	2.0	–1	1	1	1	–1	–1	1	1
0	1	3.8	0	–2	1	1	0	0	–2	–2
1	1	5.7	1	1	1	1	1	1	1	1

Tabla 7.6 ANOVA desglosado, ejemplo 7.1.

FV	SC	GL	CM	F_0	Valor-p
A_L	15.64	1	15.64	335.14	0.0000
A^2	0.30	1	0.30	6.48	0.0314
B_L	3.10	1	3.10	66.43	0.0000
B^2	0.062	1	0.062	1.34	0.28
A_LB_L	0.405	1	0.405	8.68	0.0163
A_LB^2	0.0017	1	0.0017	0.036	0.8529
A^2B_L	0.282	1	0.282	6.04	0.0363
A^2B^2	0.005	1	0.005	0.107	0.7529

Una vez más enfatizamos que, en general, no es de interés desglosar las sumas de cuadrados hasta un grado de libertad, puesto que muchas veces, como se ha comentado, la información importante se puede obtener de una descomposición menos detallada y de las gráficas de efectos. Por ejemplo, si se analiza este experimento en *Statgraphics*, el ANOVA resultante descompone los efectos principales en sus partes lineal y cuadrática, pero del efecto de interacción sólo calcula la componente lineal (A_LB_L) y los otros tres componentes de manera automática los incorpora al error.

Factoriales mixtos

Factorial mixto
Es cuando los factores en el experimento no tienen el mismo número de niveles.

Se tiene un diseño *factorial mixto* cuando los factores estudiados no tienen el mismo número de niveles. Por ejemplo, el factorial $4 \times 3 \times 2$ significa que se experimenta con tres factores, con 4, 3 y 2 niveles, respectivamente. El total de tratamientos es 24. La necesidad de utilizar un diseño factorial mixto surge por las mismas razones de un factorial 3^k (véase sección anterior). La diferencia es que el diseño factorial mixto es más frecuente que se utilice cuando, por su naturaleza discreta o categórica, los factores tienen un número finito y distinto de niveles, y el interés es estudiar todos los niveles. Por ejemplo las tres marcas de cierto material. La otra razón, aunque menos frecuente en los diseños mixtos que en los factoriales 3^k, es la posibilidad de estudiar efectos de curvatura de los factores con más de dos niveles.

En el capítulo 5 vimos una introducción a los diseños factoriales, cuyos conceptos se aplican directamente a los factoriales mixtos. Sólo que en el capítulo 5 nos limitamos a estudiar ANOVA en donde no se desglosaba un efecto en su parte lineal y cuadrática. Con las ideas vistas en la sección anterior, el análisis del diseño factorial mixto se presenta a través del siguiente ejemplo.

Ejemplo 7.2

(Continuación del ejemplo 5.2). En el ejemplo 5.2 del capítulo 5 se analiza un experimento con el que se estudia el efecto de los factores *A*: *profundidad de corte* y *B*: *velocidad de alimentación sobre el acabado* de un metal. El análisis de varianza sin desglosar se muestra en la tabla 7.7, de donde se concluye que los tres efectos *A*, *B* y *AB* están activos, y en la representación gráfica del efecto de interacción se determinó que el mejor tratamiento es profundidad baja y velocidad baja (A^-, B^-). De las gráficas de medias (figura 5.5), se hacía la observación de que, en apariencia, la velocidad tiene un efecto de curvatura importante. En la figura 5.4 del efecto de interacción se ve que algunos de los componentes de tipo cuadrático de la interacción pueden ser relevantes. Todo esto se puede verificar de manera analítica con un ANOVA desglosado.

ANOVA desglosado
ANOVA en el que los efectos compuestos se separan en sus efectos más simples que los conforman. Por ejemplo, el efecto *A* se separa en sus partes lineal y cuadrática.

ANOVA desglosado. Reiteramos que el *ANOVA desglosado* sólo aplica para los factores que son de tipo numérico, y cuando se quiere detallar el desglose de los efectos en su componente lineal y cuadrático. Las sumas de cuadrados del ANOVA de la tabla 7.7 se pueden descomponer en sumas de cuadrados con un grado de libertad, vía los contrastes. Los coeficientes para los contrastes de este ejemplo se presentan en la tabla 7.8. Se multiplica la columna de coeficientes por la columna de total para obtener el contraste. Después se aplica la fórmula dada en (7.5) para obtener la correspondiente suma de cuadrados con un grado de libertad.

Por ejemplo, si se multiplica la columna B^2 de la tabla 7.8 por la columna del total ($Y_{ij\bullet}$) y se obtiene que:

$$\text{Contraste } B^2 = (198 + 220 + \ldots + 332) = -117$$

Luego, su suma de cuadrados está dada por:

$$SC_{B^2} = \frac{(\text{Contraste } B)^2}{n\sum\sum c_{ij}^2} = \frac{(-117)^2}{3\times 24} = 190.13$$

Tabla 7.7 ANOVA sin desglosar para el ejemplo 7.2.

FV	*SC*	*GL*	*CM*	F_0	Valor-*p*
B: veloc	3 160.5	2	1 580.25	55.02	0.0000
A: prof	2 125.10	3	708.37	24.66	0.0000
AB	557.07	6	92.84	3.23	0.0000
Error	689.33	24	28.72		
Total	6 532.0	35			

Tabla 7.8 Coeficientes para calcular los contrastes en el factorial 4 × 3, ejemplo 7.2.

A	B	Total	A_L	A^2	A^3	B_L	B^2	A_LB_L	A_LB^2	A^2B_L	A^2B^2	A^3B_L	A^3B^2
–1	–1	198	–3	1	–1	–1	1	3	–3	–1	1	1	–1
–0.33	–1	220	–1	–1	3	–1	1	1	–1	1	–1	–3	3
0.33	–1	262	1	–1	–3	–1	1	–1	1	1	–1	3	–3
1	–1	299	3	1	1	–1	1	–3	3	–1	1	–1	1
–1	0	266	–3	1	–1	0	–2	0	6	0	–2	0	2
–0.33	0	290	–1	–1	3	0	–2	0	2	0	2	0	–6
0.33	0	302	1	–1	–3	0	–2	0	–2	0	2	0	6
1	0	313	3	1	1	0	–2	0	–6	0	–2	0	–2
–1	1	299	–3	1	–1	1	1	–3	–3	1	1	–1	–1
–0.33	1	298	–1	–1	3	1	1	–1	–1	–1	–1	3	3
0.33	1	317	1	–1	–3	1	1	1	1	–1	–1	–3	–3
1	1	332	3	1	1	1	1	3	3	1	1	1	1

De la tabla 7.7 se sabe que $CM_E = 28.72$ y está basado en 24 grados de libertad. Entonces el estadístico de prueba para este componente cuadrático es $F_0 = 190.13/28.72 = 6.62$, para el cual corresponde un valor-$p = 0.044$. Se concluye que el efecto B^2 está activo. Estos cálculos se pueden hacer para todos los contrastes y el ANOVA desglosado que resulta se muestra en la tabla 7.9.

En donde se observa que están activos los efectos A_L, B_L, B^2 y A_LB_L. El hecho de que B^2 sea significativo ya se había visualizado en la gráfica de efectos (figura 5.5), donde se apreciaba que la respuesta no era lineal.

Uso de software estadístico

El diseño factorial completo 3^k y fracciones de éste se incluyen en *Statgraphics*, con la secuencia: *Special* → *Experimental Design* → *Create* → *Response surface*. Después de esto se procede de manera similar que en los diseños factoriales 2^k. La novedad en

Tabla 7.9 ANOVA desglosado para el ejemplo 7.2.

FV	SC	GL	CM	F_0	Valor-p
A_L	2 108.09	1	2 108.09	79.39	0.0000
A^2	9.00	1	9.00	0.31	0.6040
A^3	8.02	1	8.02	0.28	0.6454
B_L	2 970.37	1	2 970.37	103.41	0.0000
B^2	190.12	1	190.12	6.62	0.0439
A_LB_L	429.41	1	429.41	14.95	0.0417
A_LB^2	68.47	1	68.47	2.38	0.1523
A^2B_L	0.041	1	0.041	0.001	0.9753
A^2B^2	45.12	1	45.12	1.57	0.2355
A^3B_L	0.008	1	0.008	0.0003	0.9865
A^3B^2	14.00	1	14.00	0.49	0.4980
Error	689.34	24	28.72	1.00	0.2355
Total	6 532.00	35			

el análisis es la aparición de efectos de tipo cuadrático puro (A^2, B^2, etc.), además de los efectos lineales usuales A_L, B_L y A_LB_L. En las nuevas versiones de este software la secuencia es: *Doe* → *Design Creation* → *Create* → *Response surface*.

Por su parte, los factoriales con número mixto de niveles se encuentran en la secuencia: *Special* → *Experimental Design* → *Create* → *Multilevel factorial*. Es necesario declarar el número de niveles que tiene cada factor. Después, el análisis es muy similar al del factorial 3^k.

Tanto en los diseños factoriales 3^k como en los factoriales mixtos, el software no hace la descomposición completa de las sumas de cuadrados hasta un grado de libertad, como se mostró en los ejemplos de este capítulo. Sólo descompone los efectos principales en sus componentes lineal y cuadrático puro, y de la interacción se limita a calcular los componentes de la forma A_LB_L. Los otros componentes de la interacción los incorpora como error. Sin embargo, como se comentó antes, en las gráficas de efectos se percibe la presencia de cualquier efecto de curvatura, lo cual minimiza la posibilidad de una conclusión equivocada.

Otra posibilidad de análisis donde se puede pedir todo el desglose de efectos hasta un grado de libertad es con la opción de *Multiple regression* dentro del menú principal *Relate.* Aquí, los efectos de interés se declaran en la ventana de variables independientes (la dependiente es la variable de respuesta del experimento) con el asterisco como operador de multiplicación. Por ejemplo, para obtener un modelo que incluya el término de la interacción triple *ABC* se declara *A*B*C* en dicha ventana. Una vez que se determina un modelo de regresión adecuado se interpreta obteniendo predichos y opciones gráficas.

En Minitab, la secuencia para crear los diseños vistos en este capítulo es: *Stat* → *DOE* → *Factorial* → *Create Factorial Design*, y entonces se elige la opción *General full factorial*.

Preguntas y ejercicios

1. ¿En qué situaciones prácticas, en el contexto de un diseño factorial, se suelen utilizar tres niveles o más para un factor?
2. ¿Qué desventajas prácticas se tienen al utilizar un factor con tres o más niveles en un diseño factorial?
3. Represente de manera gráfica un efecto lineal y otro cuadrático.
4. ¿A qué tipo de diseños factoriales se les llama mixtos?
5. Se desea investigar de qué manera afecta el tiempo de curado y el tipo de acelerante a la resistencia del caucho vulcanizado. Se realiza un experimento y se obtienen los siguientes datos.

Tiempo de curado	Acelerante		
	A	*B*	*C*
40	3 900, 3 600	4 300, 3 700	3 700, 4 100
60	4 100, 3 500	4 200, 3 900	3 900, 4 000
80	4 000, 3 800	4 300, 3 600	3 600, 3 800

a) Señale el nombre del diseño de experimento utilizado y su modelo estadístico.
b) Formule todas las hipótesis que se pueden probar.
c) Realice el análisis estadístico apropiado para probar las hipótesis que formuló.
d) En caso de haberlo, señale el tiempo de cura que es mejor para aumentar la resistencia.
e) Señale el acelerante que es mejor (si es que lo hay), para aumentar la resistencia.
f) ¿Hay alguna combinación de tiempo y acelerante que sea mejor? Diga cuál es, si la hay.
g) Explique en forma esquemática cómo se calcula el valor-*p* para el tiempo de cura.
h) Verifique que se cumplan los supuestos. En caso de que no se cumpliera el supuesto de igual varianza para tiempo de cura, ¿qué significaría eso?

6. Una planta química produce oxígeno. La pureza del oxígeno depende de la *temperatura* y la *presión*. Las condiciones de operación usuales son $X_1 = -220°C$ y $X_2 = 1.3$. Interesa encontrar mejores condiciones de operación, para lo cual se obtuvieron los siguientes datos:

X_1	X_2	Pureza
–225	1.1	82.8
–225	1.3	85.5
–215	1.1	84.7
–215	1.3	86.0
–220	1.2	84.2
–220	1.2	84.5
–220	1.2	83.9
–220	1.2	84.3

a) ¿Qué diseño es éste?
b) Analice los datos y obtenga conclusiones.

7. Se cree que la adhesividad de un pegamento depende de la presión y de la temperatura al ser aplicado. Se realiza un experimento factorial con ambos factores fijos.

	Temperatura (°F)		
Presión (lb/pulg²)	250	260	270
120	9.60	11.28	9.00
130	9.69	10.10	9.57
140	8.43	11.01	9.03
150	9.98	10.44	9.80

a) Formule las hipótesis y el modelo estadístico que se desea probar.
b) Analice los datos y obtenga las conclusiones apropiadas.
c) ¿Se puede analizar si hay interacción entre los dos factores controlados?
d) Verifique los supuestos.

8. Se desea investigar el efecto del tipo y cantidad (%) de almidón sobre la friabilidad (%) de tabletas. Se corre un diseño 3×3 con $n = 4$ réplicas. Los datos obtenidos se muestran a continuación:

	Almidón (%)		
Tipo de almidón	26%	30%	40%
Pregelatinizado	0.7782	0.4272	0.7192
	0.7654	0.4336	0.6742
	0.7592	0.4552	0.6892
	0.7758	0.4771	0.7023
Almidón	0.7856	0.6273	0.9562
	0.8093	0.6592	0.9656
	0.8126	0.6692	0.9656
	0.8172	0.6523	0.9231
Dextrina	0.8543	0.8023	1.1356
	0.8792	0.7986	1.1923
	0.8723	0.7992	1.1643
	0.8993	0.7827	1.1732

a) Escriba el modelo estadístico más apropiado para el diseño.
b) Obtenga el ANOVA sin desglosar y obtenga conclusiones.
c) Realice la gráfica de efectos principales y de interacción, y destaque los aspectos más relevantes.
d) De la gráfica de efectos principales para el factor % de almidón, ¿hay algún tipo de evidencia de que el efecto no sea lineal? Argumente su respuesta.
e) Verifique supuestos a través de gráficas de residuales.
f) Obtenga el ANOVA desglosado; para ello, el efecto lineal y cuadrático debe desglosarse sólo para el factor % de almidón, ya que es el único que es cuantitativo. Comente lo obtenido y contrástelo con lo observado en los incisos *c*) y *d*).

9. La mata ratón (*Gliricidia sepium*) es de interés económico para fines forrajeros. Para propagar su uso por parte de los ganaderos es importante investigar la forma adecuada de sembrar esta especie. Por ello, se plantea un experimento factorial (Contreras y Ochoa, 2003). Los factores de control son la longitud y el diámetro de las estacas, mientras que la variable de interés es el porcentaje de estacas que rebrotan. Se utilizaron tres niveles para ambos factores:

Factor	Niveles
Longitud	$L_1 = 10$, $L_2 = 15$ y $L_3 = 20$ cm
Diámetro	$D_1 = 2.5\text{-}3.5$, $D_2 = 3.6\text{-}4.5$, $D_3 = 4.6\text{-}5.5$ cm

Se sembraron 4 estacas de cada combinación de longitud y diámetro. Los resultados obtenidos en porcentaje de estacas rebrotadas después de 42 días se muestran a continuación:

Longitud	Diámetro	Estacas con brotes (%)
L_1	D_1	0
L_1	D_2	0
L_1	D_3	75
L_2	D_1	0
L_2	D_2	25
L_2	D_3	75
L_3	D_1	0
L_3	D_2	0
L_3	D_3	100

a) Analice el experimento usando el ANOVA sin desglosar y desglosado.
b) ¿Cuál de los dos factores es el más crítico para observar brotes?
c) ¿Qué comentario corresponde a las pocas estacas sembradas en cada tratamiento?
d) ¿Cuáles son las dimensiones recomendadas para observar más brotes? Grafique los efectos estimados.

10. Como resultado del experimento anterior se generó la hipótesis de que "las estacas cortas, con longitudes entre 8 y 10 cm y con diámetros mayores a 5 cm pudieran tener una buena capacidad de rebrote". Para confirmarla se plantea un segundo experimento con 10 estacas; cada combinación de las longitudes $L_1 = 8$ cm y $L_2 = 10$ cm con los diámetros $D_1 =$ 5-6 cm, $D_2 =$ 6.1-7 cm y $D_3 =$ 7.1-8 cm. Después de 65 días se midieron las respuestas $Y_1 =$ porcentaje de estacas con brotes, $Y_2 =$ rebrotes por estaca y $Y_3 =$ altura de las plántulas. Los resultados son los siguientes:

Longitud	Diámetros	Y_1	Y_2	Y_3
L_1	D_1	60	0.9	2.80
L_1	D_2	60	0.7	2.35
L_1	D_3	70	0.9	3.60
L_2	D_1	90	1.2	5.10
L_2	D_2	100	1.6	7.30
L_2	D_3	100	2.1	7.45

a) ¿Considera razonable la hipótesis generada a partir del primer experimento?
b) ¿Cree que la conjetura se confirma con el segundo estudio? Haga un ANOVA sin desglosar y desglosado a un grado libertad para analizar Y_1.
c) Analice las respuestas Y_2 y Y_3. ¿Con estas respuestas resulta significativo el factor diámetro? Explique y pruebe las hipótesis correspondientes.
d) ¿Cuáles serían las dimensiones más recomendadas para las estacas?

11. Se ha observado que lo esponjoso de los *hot-cakes* (o panqueques) puede variar bastante dependiendo de dónde se compren o cómo se hagan. Arroyo-Gutiérrez, *et al*. (2003) reportan un experimento donde investigan lo esponjoso de los *hot-cakes* en

función de la marca de harina comercial: Great Value (GV), Gamesa (GA) y Jamima (JA), y del tipo de leche: Parmalat light (PL), Parmalat entera (PE) y Carnation evaporada (CC). En el orden mencionado, las harinas cubren un rango de precios de 9 a 19 pesos y las leches van desde 2 hasta 48 gramos de grasa por litro. El objetivo del experimento es obtener los *hot cakes* más esponjosos o gruesos en su centro, medido esto con un palillo de madera sobre un plato plano. Se aleatorizó el orden de corrida para las nueve combinaciones. En cada caso se hizo mezcla suficiente para preparar cuatro *hot-cakes* (que con reservas se puede ver como réplica). En cada mezcla se agregó margarina y un huevo, elegido al azar por aquello de su tamaño. Los grosores de los *hot-cakes* se muestran a continuación:

Tratamiento	Harina	Leche	Rep 1	Rep 2	Rep 3	Rep 4
1	GV	PL	9	8	9	9
2	GV	PE	12	12	10	9
3	GV	CC	15	14	15	13
4	GA	PL	16	14	19	17
5	GA	PE	13	14	13	12
6	GA	CC	19	18	21	18
7	JA	PL	17	19	17	18
8	JA	PE	13	15	12	11
9	JA	CC	16	16	15	20

a) ¿Que implicaciones puede tener el hecho de que las repeticiones sean *hot-cakes* hechos con la misma harina que se preparó una sola vez?
b) Mencione algunos otros factores que pudieron influir en los resultados y que, por lo tanto, tuvieron que mantenerse fijos.
c) Obtenga el ANOVA sin desglosar (suponiendo que las repeticiones son auténticas) y las gráficas de los efectos.
d) Interprete los resultados. ¿Cuál tratamiento produce los *hot-cakes* más esponjosos?
e) Verifique los supuestos de normalidad y varianza constante.

12. Se quiere estudiar la forma de bajar la presión sanguínea diastólica en función de varias combinaciones de fármacos y dietas (Langford *et al*., 1991). La variable de interés es el cambio promedio en la presión sanguínea después de seis meses que se midió la presión de referencia a los participantes en el estudio. En cada combinación de fármaco y dieta se tienen entre 80 y 90 personas. Los tres fármacos fueron un placebo, chlorthalidone (25 mg) y atenolol (50 mg), y las dietas fueron: la usual, una de pérdida de peso y una tercera baja en sodio y alta en potasio. Los datos obtenidos se muestran en la siguiente tabla, donde el signo negativo indica que la presión promedio bajó después de seis meses y la segunda variable de respuesta es el porcentaje de personas con presión diastólica debajo de 90 mm Hg.

Tratamiento	Cambio de Presión	% debajo de mm Hg
F_1, D_1	–7.96	71.1
F_2, D_1	–10.78	86.2
F_3, D_1	–12.43	93.1
F_1, D_2	–8.78	73.1
F_2, D_2	–15.06	93.1
F_3, D_2	–14.81	92.0
F_1, D_3	–7.91	64.1
F_2, D_3	–12.18	86.5
F_3, D_3	–12.76	87.8

a) Analice las dos respuestas con los ANOVA más apropiados.
b) ¿Hay evidencia suficiente para afirmar que, además del fármaco, la dieta contribuye de manera significativa a bajar la presión sanguínea? Argumente su respuesta.
c) ¿Cuál tratamiento es el mejor?

13. Interesa estudiar el rendimiento de plantas de tres selecciones A-16R, A-8R y A-7R de chile dulce (pimiento), en función de la edad al momento del transplante (Montaño-Mata y Núñez, 2003). Las variables de respuesta fueron el rendimiento en toneladas por hectárea (Y_1) y el número promedio de frutos por planta (Y_2). El experimento y los datos obtenidos se muestran en la siguiente tabla, donde el primer número de cada casilla corresponde a Y_1 y el segundo a Y_2.

Edad (días)	Respuesta	Selección		
		A-7R	A-8R	A-16R
50	Y1 Y2	13.00 88.00	16.07 115.33	12.13 83.67
45	Y1 Y2	15.00 106.67	13.95 101.33	17.87 119.33
40	Y1 Y2	15.06 105.33	13.89 108.00	13.08 87.67
35	Y1 Y2	10.77 78.67	10.72 83.33	10.45 73.00

a) Escriba los modelos utilizados y obtenga el mejor ANOVA sin desglosar para ambas variables de respuesta.
b) ¿Cómo influye la edad, al momento del transplante, en el rendimiento de las plantas?
c) ¿El número de frutos promedio por planta tiene relación directa con el rendimiento?
d) ¿Cuál es el mejor tratamiento?
e) Verifique los supuestos usuales.

14. Se realiza un experimento para comparar la proporción de palomitas de maíz que se forman (% de granos que reventaron) con tres marcas de palomitas para horno de microondas. Se utilizan hornos con dos potencias diferentes y tres tiempos de permanencia en el horno, con 2 réplicas. Se obtienen los siguientes resultados:

Marca	Potencia	Tiempo		
		4 min	4.5 min	5 min
1	500	73.8, 65.5	72.7, 81.9	70.3, 91.0
1	625	70.8, 75.3	74.1, 72.1	78.7, 88.7
2	500	45.3, 47.6	73.7, 65.8	93.4, 76.3
2	625	66.3, 45.7	79.3, 86.5	92.2, 84.7
3	500	51.4, 67.7	62.5, 65.0	50.1, 81.5
3	625	64.0, 77.0	71.5, 80.0	82.1, 74.5

a) ¿Qué diseño se utilizó? Escriba el modelo estadístico correspondiente.
b) Analice estos datos y obtenga conclusiones.
c) Grafique los efectos significativos e interprételos.
d) ¿Cuál es la potencia del horno y el tiempo recomendados para cada marca de palomitas?
e) ¿Cuál es el mejor tratamiento considerando los tres factores, y cuál es el porcentaje de granos reventados que se esperarían en tal tratamiento?
f) ¿Cuál de las marcas de palomitas se afecta menos (más robusta) debido a la acción del horno y el tiempo?
g) Compruebe los supuestos de normalidad y varianza constante.

Capítulo 8
Diseños factoriales fraccionados 2^{k-p}

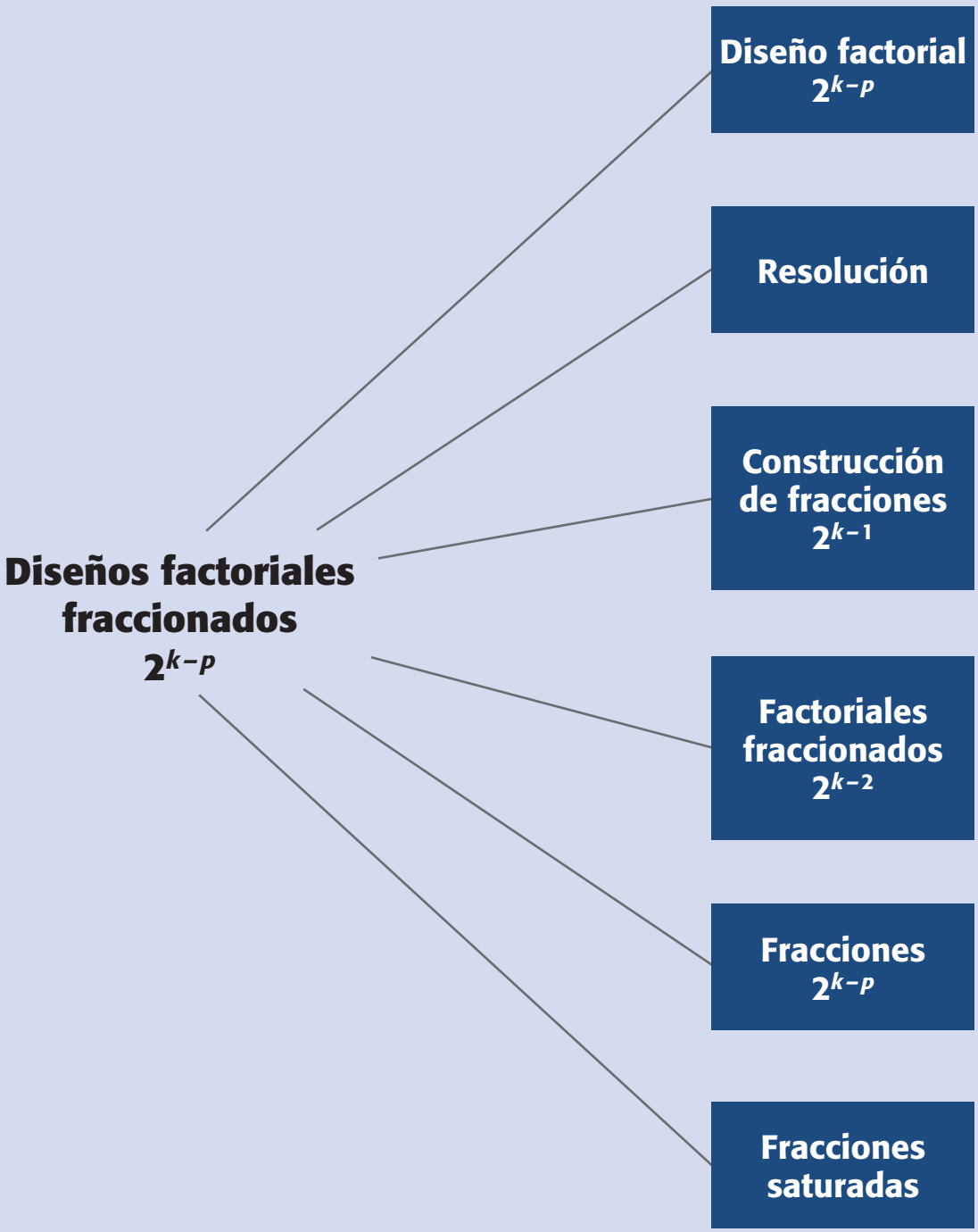
Sumario

- Diseño factorial fraccionado 2^{k-1}
- El concepto de resolución
- Construcción de fracciones 2^{k-1}
- Experimento 2^{5-1}: ejemplo integrador
- Diseños factoriales fraccionados 2^{k-2}
- Diseño factorial fraccionado 2^{k-p}
- Experimento 2^{7-4}: ejemplo integrador
- Tópicos adicionales sobre factoriales fraccionados
- Uso de software

Objetivos de aprendizaje

- Conocer los aspectos principales de los diseños factoriales y saber cómo y cuándo aplicarlos.
- Comprender los conceptos de resolución III, IV y V, así como su aplicación en la elección de una fracción apropiada.
- Construir fracciones a cualquier grado de fraccionamiento (2^{k-p}).
- Seleccionar la fracción adicional más adecuada para aclarar ambigüedades heredadas de una primera fracción.

Mapa conceptual



Conceptos clave

- Diseño de Plackett-Burman
- Diseño de resolución III
- Diseño de resolución IV
- Diseño de resolución V
- Diseño factorial fraccionado 2^{k-p}
- Diseños factoriales fraccionados
- Efectos alias
- Efectos generadores
- Estructura del alias
- Factorial fraccionado saturado
- Fracción 2^{k-1}
- Generador de la fracción
- Mínima aberración
- Relación definidora
- Resolución

Diseños factoriales fraccionados

Diseños en los que se elige adecuadamente una parte o fracción de los tratamientos de un factorial completo, con la intención de estudiar el efecto de los factores utilizando menos corridas experimentales.

Cuando crece el número de factores también aumenta rápidamente el número de tratamientos en los diseños factoriales completos 2^k. Por ejemplo, para $k = 6$ factores, una sola réplica del diseño factorial completo 2^6 implica correr 64 pruebas, que corresponden al número de tratamientos del diseño; para $k = 7$ son $2^7 = 128$ puntos de diseño. En la práctica no es posible hacer tantas corridas experimentales. Sin embargo, es frecuente que en las primeras etapas de una investigación interese estudiar muchos factores, digamos 6 o más. Para experimentar con esta cantidad de factores se requiere una estrategia que permita reducir de manera importante el número de tratamientos experimentales, pero que al mismo tiempo se pierda el mínimo de información valiosa. Tal estrategia la conforman los *diseños factoriales fraccionados*, los cuales, gracias al exceso de información que acumulan los diseños factoriales completos cuando se estudian muchos factores, permiten sacrificar información poco importante en aras de un diseño manejable en cuanto al número de corridas experimentales. Las corridas en los factoriales fraccionados son una parte o una fracción de los tratamientos de los factoriales completos. La teoría de diseños fraccionados se basa en una jerarquización de los efectos: son más importantes los efectos principales, seguidos por las interacciones dobles, luego las triples, cuádruples, etcétera.

En la tabla 8.1 se muestra el número de efectos potencialmente de mayor interés para diferentes diseños factoriales 2^k, de ahí se observa que el primer diseño factorial completo que genera un importante exceso de información es el factorial completo 2^5, ya que éste permite estimar 31 efectos, de los cuales sólo 15 son potencialmente importantes (los 5 efectos principales más las 10 interacciones dobles) y los 16 efectos restantes, conformados por las interacciones de tres o más factores, prácticamente se pueden ignorar *a priori*. Más adelante se muestra que con una fracción a la mitad del diseño factorial completo 2^5 ($\frac{1}{2}\ 2^5 = 2^{5-1}$), se pueden estimar limpiamente los 15 efectos potenciales importantes, sacrificando la información relativa a las 16 interacciones de alto orden que no interesan. Así, con un diseño factorial fraccionado 2^{5-1} se obtiene esencialmente la misma información que con el factorial completo 2^5, pero con la mitad del costo experimental.

En la tabla 8.1 también se observa que para menos de cinco factores ($k < 5$) los efectos potencialmente importantes superan en número a los efectos ignorables *a priori*, de aquí que si se fraccionan estos diseños, es forzoso que se pierda información que puede ser relevante. Por otro lado, cuando $k \geq 5$ el número de efectos ignorables supera al número de efectos no ignorables o potencialmente importantes, lo cual indica que estos diseños se pueden fraccionar sin perder información valiosa.

Tabla 8.1 Efectos en los factoriales 2^k.

	Diseño 2^k	Total de efectos	Efectos no ignorables	Efectos ignorables
	2^2	3	3	0
	2^3	7	6	1
	2^4	15	10	5
⇒	2^5	31	15	16
	2^6	63	21	42
	2^7	127	28	99

Mientras más grande es el valor de k, el diseño admite un grado de fraccionamiento mayor.

Cabe aclarar que al *correr* sólo una fracción del diseño factorial completo ocurren dos hechos insoslayables:

1. Se pierde información, ya que habrá efectos que no podrán estimarse y se tienen menos grados de libertad disponibles para el error. Los efectos que se pierden se espera que sean, en la medida de lo posible, interacciones de alto orden, las cuales se pueden ignorar de antemano con bajo riesgo.
2. Los efectos que sí se pueden estimar tienen al menos un *alias*. El que un *efecto* sea *alias* de otro significa que en realidad son el mismo efecto con nombres distintos, y al estimar a uno de ellos al mismo tiempo se estima el otro, de manera que no se pueden separar. Cuando el experimentador elige una fracción en la que dos efectos potencialmente importantes son alias, debe contar de antemano con una estrategia de interpretación del efecto estimado.

Efectos alias
Son dos o más efectos con nombres diferentes que comparten el *mismo* contraste y, por lo tanto, estiman el mismo efecto.

Estos dos hechos se entenderán mejor en el contexto de las fracciones que se construyan en los ejemplos de este capítulo.

Diseño factorial fraccionado 2^{k-1}

Fracción 2^{k-1}
La mitad de los tratamientos del factorial completo 2^k: $\frac{1}{2}\ 2^k = 2^{k-1}$.

La *notación* 2^{k-1} significa una *fracción a la mitad* del diseño factorial completo 2^k, $k > 2$ con $\frac{1}{2}\ 2^k = 2^{k-1}$. No tiene sentido fraccionar el diseño factoria1 2^2 porque prácticamente desaparece: al tener sólo cuatro tratamientos, fraccionarlo a la mitad implicaría correr dos tratamientos y con ellos no se podrían estimar ni siquiera los dos efectos principales. A continuación se muestra cómo se fracciona a la mitad un diseño factorial y se ilustra el método para 3, 4 y 5 factores.

Diseño factorial fraccionado 2^{3-1}

El primer diseño que se puede fraccionar (aunque veremos que no se recomienda hacerlo) es el factorial completo 2^3, el cual, escrito en la notación estándar, se muestra en la tabla 8.2. Si queremos fraccionar a la mitad este diseño, entonces es necesario seleccionar cuatro de entre los ocho tratamientos. De entrada sabemos que existen

Tabla 8.2 Diseño factorial completo 2^3 y contraste *ABC*.

A	B	C		ABC
–1	–1	–1		–
1	–1	–1		+
–1	1	–1		+
1	1	–1	⇒	–
–1	–1	1		+
1	–1	1		–
–1	1	1		–
1	1	1		+

$\binom{8}{4} = 70$ posibles grupos de 4 tratamientos, por lo que surge la interrogante sobre cuál o cuáles de esas 70 posibles fracciones son las más apropiadas. Es adecuada aquella fracción (diseño factorial fraccionado 2^{3-1}) que permita estimar los efectos de mayor interés.

Recordemos que con el diseño 2^3 completo se pueden estimar siete efectos: A, B, C, AB, AC, BC y ABC. De acuerdo con su jerarquía, el efecto menos importante *a priori* es la interacción triple ABC, así que éste es el efecto más sacrificable para generar la fracción a la mitad, de manera que se pierda un mínimo de información. La generación de la fracción se hace con base en los signos del contraste ABC: los signos "+" del contraste ABC señalan a los tratamientos que conforman la llamada *fracción principal*, y los signos "–" indican la *fracción complementaria* (véase tabla 8.2). Los dos diseños factoriales fraccionados 2^{3-1} así generados, proporcionan la misma calidad de información sobre los efectos potencialmente importantes. Se puede mostrar que cualquier otra elección de cuatro tratamientos daría peores resultados. En la tabla 8.3, la fracción 1 es la fracción principal generada por $I = +ABC$ y la fracción 2 o complementaria se genera con $I = -ABC$. La letra I surge porque este efecto generador hace las veces de identidad o neutro multiplicativo, como se muestra más adelante. De aquí que el efecto no estimable ABC se llama *generador de la fracción*, puesto que su contraste es la base para construir las dos fracciones.

Generador de la fracción
Efecto cuyo contraste es utilizado para generar la fracción factorial. Este efecto no se puede estimar con esa fracción.

Obsérvese que en las corridas que forman ambas fracciones en la tabla 8.3, todos los factores están dos veces en nivel más y dos veces en nivel menos, además (por ejemplo los dos primeros tratamientos de la fracción 1) en las corridas en que un factor tiene los mismos signos, los otros factores tienen un signo más y uno menos. Lo anterior se desprende de la *propiedad de ortogonalidad*. Al correr alguna de las fracciones que se muestran en la tabla 8.3 no se podrá estimar el *efecto ABC*, puesto que no tiene contraste; por ejemplo, para la fracción 1, el *contraste ABC* tiene sólo signos +, por lo que no existe tal contraste. Debido a que el "*contraste ABC*" sólo tiene signos positivos, podemos decir que se confunde o se alía con el total de los datos, o dicho de otro modo, el *efecto ABC* se confunde con la media global μ.

En la figura 8.1 se muestra la representación geométrica de las dos fracciones del diseño 2^3, nótese que cada fracción tiende a cubrir toda la región experimental delimitada por el cubo.

Estructura de alias del diseño 2^{3-1} con $I = ABC$. Al estimar los efectos potencialmente importantes con cualquiera de las fracciones dadas en la tabla 8.3, resulta que cada efecto estimado tiene un alias. Consideremos, por ejemplo, la fracción 1 de la tabla 8.3. Este diseño se generó con $I = +ABC$, que en este caso también es

Tabla 8.3 Dos posibles diseños fraccionados 2^{3-1}.

Fracción 1 ($I = +ABC$)				Fracción 2 ($I = -ABC$)			
A	B	C		A	B	C	
1	–1	–1	a	–1	–1	–1	(1)
–1	1	–1	b	1	1	–1	ab
–1	–1	1	c	1	–1	1	ac
1	1	1	abc	–1	1	1	bc

la *relación definidora*, ya que define totalmente la *estructura del alias* del diseño, la cual consiste en escribir de manera explícita cuáles son los alias de cada efecto, y esta estructura se deduce fácilmente del generador de la fracción, considerando el signo utilizado. Por ejemplo, en la fracción 1 de la tabla 8.3, el contraste del efecto *A* está dado por:

$$Contraste\ A = (a + abc - b - c)$$

mientras que al multiplicar las columnas $B \times C$ se obtiene,

$$Contraste\ BC = (a + abc - b - c)$$

Observe que *Contraste A* = *Contraste BC*, lo cual significa que los efectos *A* y *BC* son alias. Al estimar el *efecto A* también se estima el *efecto BC*. Así, en realidad se estima la suma *A* + *BC* de ambos efectos y no se sabe con certeza cuál es el que predomina o si ambos afectan. De igual forma se puede ver que en la fracción 1:

$$Contraste\ B = Contraste\ AC = (b + abc - a - c)$$

$$Contraste\ C = Contraste\ AB = (c + abc - b - a)$$

así que *B* es alias de *AC*, y *C* es alias de *AB*. En resumen, la estructura del alias del diseño factorial fraccionado 2^{3-1} está dada por:

$$\begin{array}{l} A + BC \\ B + AC \\ C + AB \end{array} \qquad \textbf{(8.1)}$$

Otra forma de obtener la *estructura alias* de un diseño factorial fraccionado es con la relación definidora del diseño. Al multiplicar cada efecto por esta relación, con el uso de *multiplicación módulo* 2: que significa que al multiplicar cualquier efecto por la identidad es igual al efecto, y al multiplicar un efecto por sí mismo es igual a la identidad. Por ejemplo, si aplicamos esto para el efecto *A*, tendríamos que $A \times I = A$ y que $A \times A(Mód2) = A^2 = A^0 = I$. Entonces, el alias de *A* se obtiene al multiplicar por *A* ambos lados de la relación $I = ABC$, y resulta que,

$$A \times I = A \times ABC = A^2BC = BC$$

Los alias de *B* y C se obtienen con:

$$B \times I = B \times ABC = AB^2C = AC$$

$$C \times I = C \times ABC = ABC^2 = AB$$

con lo que resulta la estructura de alias que obtuvimos antes, a partir de los contrastes. Cuando se fracciona un diseño más grande resulta más práctico obtener la estructura de alias a partir de la relación definidora del diseño que de los contrastes.

Relación definidora
Está dada por los generadores de una fracción, más todos sus posibles productos entre sí módulo 2. Si hay un generador éste también es la relación definidora.

Estructura del alias
En ella se definen de manera explícita los alias de cada efecto.

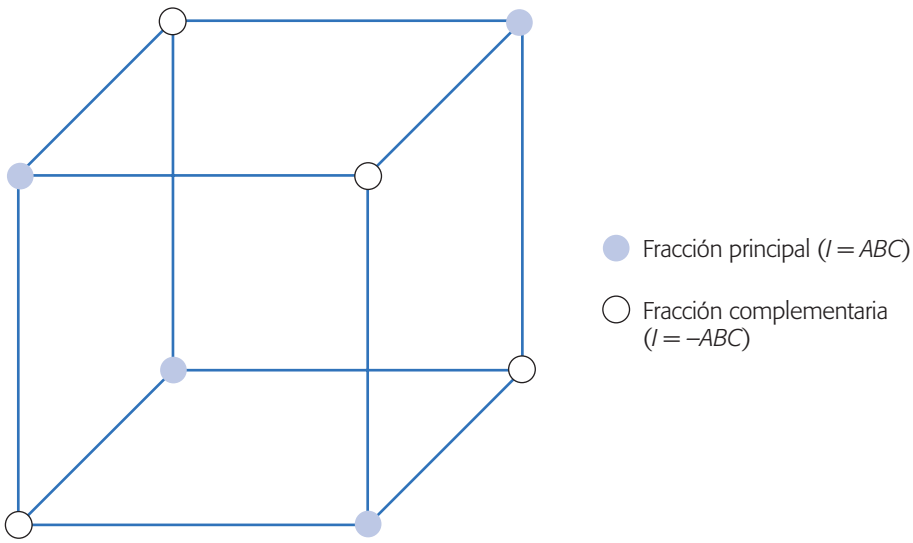


Figura 8.1 Representación de los diseños factoriales fraccionados 2^{3-1}.

Note que han ocurrido los dos hechos que se mencionan en la introducción: por construcción de la fracción se perdió la información relativa a la interacción triple *ABC* y cada efecto estimable tiene al menos un alias, lo que en ocasiones puede entorpecer su interpretación.

Interpretación de efectos alias. Para interpretar los efectos alias o aliados es necesario suponer que sólo uno de ellos es responsable del efecto observado y que los demás efectos son nulos. Por ejemplo, si se utilizara el diseño 2^{3-1} con la estructura de alias dada en la fórmula (8.1), los efectos alias se interpretan atribuyendo (por jerarquía) el efecto observado al efecto principal de cada grupo y considerando nulas las interacciones dobles. Comúnmente resulta bastante riesgoso suponer *a priori* que las interacciones dobles no afectan, de aquí que no se recomienda utilizar este diseño, a menos que el costo experimental no permita correr el factorial completo. En general no es buena estrategia utilizar diseños fraccionados donde se alían dos efectos que son potencialmente importantes, como son los efectos principales y las interacciones dobles; sin embargo, habrá situaciones en las que no queda otra alternativa.

Estructura de alias del diseño 2^{3-1} con $I = -ABC$. La estructura alias para el diseño 2^{3-1} con relación definidora $I = -ABC$ (fracción 2, tabla 8.3) está dada por:

$$\begin{aligned} &A - BC \\ &B - AC \\ &C - AB \end{aligned} \qquad \textbf{(8.2)}$$

En este caso, al estimar los efectos *A*, *B* y *C* realmente se estiman $A - BC$, $B - AC$ y $C - AB$, respectivamente.

El concepto de resolución

Al correr un diseño factorial fraccionado los efectos no pueden estimarse de manera aislada, sino que se estiman las sumas (o restas) de efectos alias. La interpretación de aliados que se suman se hace fácilmente si puede suponerse que todos los sumandos, excepto uno, no son importantes. Así, el efecto total se puede atribuir a este único efecto que se considera relevante. Entonces, la estrategia es elegir, siempre que sea posible, diseños fraccionados en los cuales los efectos potencialmente importantes sean alias de efectos que sea razonable suponer que son poco importantes. Esto se logra con diseños factoriales fraccionados que tengan alta resolución. La *resolución* es una característica de un factorial fraccionado, que indica qué tan bien pueden estudiarse los efectos potencialmente importantes mediante tal diseño.

Las resoluciones de mayor interés son las que se describen a continuación:

1. *Diseños de resolución III*. En estos diseños los efectos principales no son alias entre ellos, pero existen efectos principales que son alias de alguna interacción doble. Por ejemplo, el diseño 2^{3-1} con relación definidora $I = ABC$ (o $I = -ABC$) es de resolución III.
2. *Diseños de resolución IV*. En este diseño los efectos principales no están alias entre ellos ni con las interacciones dobles, pero algunas interacciones dobles están alias con otra interacción doble. Por ejemplo, el diseño 2^{4-1} con relación definidora $I = ABCD$ (o $I = -ABCD$) es de resolución IV.
3. *Diseños de resolución V*. En estos diseños los efectos principales y las interacciones dobles están alias con interacciones triples o de mayor orden, es decir, los efectos principales e interacciones dobles están limpiamente estimados. Por ejemplo, el diseño 2^{5-1} con relación definidora $I = ABCDE$ (o $I = -ABCDE$) es de resolución V.

Resolución
Es una característica de un factorial fraccionado, que indica qué tan bien pueden estudiarse los efectos potencialmente importantes mediante tal diseño.

Diseño de resolución III
Es cuando los efectos principales no son alias entre ellos, pero existen efectos principales que son alias de interacciones dobles.

Diseño de resolución IV
Cuando los efectos principales no tienen alias entre ellos ni con las interacciones dobles, pero algunas interacciones dobles son alias entre ellas.

Diseño de resolución V
Es cuando los efectos principales y las interacciones dobles están alias con interacciones triples o de mayor orden.

Una definición de resolución es la siguiente: Un diseño factorial fraccionado es de resolución R si los efectos formados por la interacción de P factores, no son alias de efectos de interacción que tengan menos de $R - P$ factores.

En general, en los diseños factoriales fraccionados en dos niveles, la resolución está dada por la "palabra o efecto" de la relación definidora con el menor número de letras. Por ejemplo, en los diseños 2^{k-1} la resolución es igual al número de letras del generador, ya que al mismo tiempo éste es la relación definidora. Así, las fracciones 2^{3-1}, 2^{4-1} y 2^{5-1} tienen resolución III, IV y V, respectivamente, porque sus generadores correspondientes se componen de 3, 4 y 5 letras.

Construcción de fracciones 2^{k-1}

Una manera de construir en dos pasos diseños fraccionados 2^{k-1} con la más alta resolución posible es la siguiente:

1. Se lista el diseño factorial completo para $k - 1$ factores, y de esta forma se tienen las primeras $k - 1$ columnas de la fracción deseada.

2. La columna faltante (la k-ésima) se construye multiplicando entre sí las columnas previas. Si se quiere la fracción complementaria se cambian los signos de esta última columna. El diseño que resulta es un diseño factorial fraccionado 2^{k-1} con resolución máxima $R = k$.

Ejemplo 8.1

Construcción del diseño 2^{4-1}. Estos dos pasos aplicados en la construcción del diseño factorial fraccionado 2^{4-1} con resolución IV y con generador $I = -ABCD$ quedan de la siguiente manera:

1. Primero se lista el diseño factorial completo $2^{4-1} = 2^3$ dado por:

A	B	C	D
–	–	–	
+	–	–	
–	+	–	
+	+	–	
–	–	+	
+	–	+	
–	+	+	
+	+	+	

dejando en blanco los espacios para los niveles del factor D.

2. La columna faltante de niveles para el factor D se obtiene al multiplicar las columnas A, B y C de acuerdo al generador. En este caso el generador indica que $D = -ABC$. Haciendo el producto $-ABC$ se obtienen los signos de la cuarta columna:

A	B	C	D = –ABC
–	–	–	+
+	–	–	–
–	+	–	–
+	+	–	+
–	–	+	–
+	–	+	+
–	+	+	+
+	+	+	–

Si se quisiera la fracción principal que tiene generador $I = +ABC$, el primer paso es el mismo, y en el segundo paso los niveles de D se obtienen con el producto positivo de las columnas ($D = +ABC$). Una ventaja de la fracción complementaria que se acaba de construir con respecto a la fracción principal, es que no contiene las combinaciones de niveles más extremosas (–, –, –, –) y (+, +, +, +).

Experimento 2^{5-1}: ejemplo integrador

Este ejemplo ya fue presentado en la sección "Experimento 2^5 no replicado: ejemplo integrador" del capítulo 6, como un diseño factorial completo 2^5 con una sola réplica.

Ahora mostraremos que si se hubiera corrido sólo una fracción a la mitad (2^{5-1}), las conclusiones hubieran sido las mismas que se obtuvieron con el factorial completo, con la diferencia de que el costo experimental hubiera sido la mitad. El objetivo de este ejemplo es mostrar que al correr una fracción a la mitad de un diseño factorial completo, con cinco factores, no se pierde (necesariamente) información valiosa y se obtienen las mismas conclusiones.

Recordemos que el experimento se desarrolló en una planta donde se fabrican semiconductores, en la cual se quería mejorar el rendimiento del proceso mediante el diseño de experimentos. De acuerdo con la experiencia del grupo de mejora, los factores que podían tener mayor influencia sobre la variable de respuesta (rendimiento), así como los niveles de prueba utilizados, fueron los siguientes:

A = Nivel de la abertura (pequeña, grande).
B = Tiempo de exposición (20% abajo, 20% arriba).
C = Tiempo de revelado (30 s, 45 s).
D = Dimensión de la máscara (pequeña, grande).
E = Tiempo de grabado (14.5 min, 15.5 min).

Consideremos las fracciones a la mitad para este diseño 2^5. La construcción de las fracciones se hace en dos pasos: 1) Se escribe el diseño 2^{5-1} como si fuera un factorial completo 2^4. 2) La columna faltante es el producto de las cuatro primeras. Esta fracción corresponde a la fracción principal (generador $I = ABCDE$). Si se toma el producto negativo de las cuatro columnas se obtiene la fracción complementaria (generador $I = -ABCDE$). Ambas fracciones juntas conforman el diseño factorial completo 2^5. Las dos fracciones con sus respectivos datos se muestran en la tabla 8.4.

Análisis de las dos fracciones

En la figura 8.2 se muestran los diagramas de Pareto para los efectos de cada una de las fracciones. Vemos que cualquiera de las fracciones detecta los efectos *A*, *B*, *C* y *AB* como los más importantes. Estos paretos son básicamente iguales que en el análisis del diseño completo (véase capítulo 6).

Con los gráficos de Daniel pasa lo mismo: ambas fracciones detectan los mismos efectos, que en apariencia son significativos (véase figura 8.3).

Baste decir que los análisis de varianza resultan los mismos, debido a la claridad con la que trabajan los paretos y los gráficos de Daniel.

Tabla 8.4 Dos fracciones 2^{5-1}.

Fracción principal		Fracción complementaria	
e = 8.0	*d* = 8.0	(1) = 7.0	*de* = 6.0
a = 9.0	*ade* = 10.0	*ae* = 12.0	*ad* = 10.0
b = 34.0	*bde* = 30.0	*be* = 35.0	*bd* = 32.0
abe = 52.0	*abd* = 50.0	*ab* = 55.0	*abde* = 53.0
c = 16.0	*cde* = 15.0	*ce* = 15.0	*cd* = 18.0
ace = 22.0	*acd* = 21.0	*ac* = 20.0	*acde* = 20.0
bce = 45.0	*bcd* = 44.0	*bc* = 40.0	*bcde* = 41.0
abc = 60	*abcde* = 63.0	*abce* = 65.0	*abcd* = 61.0

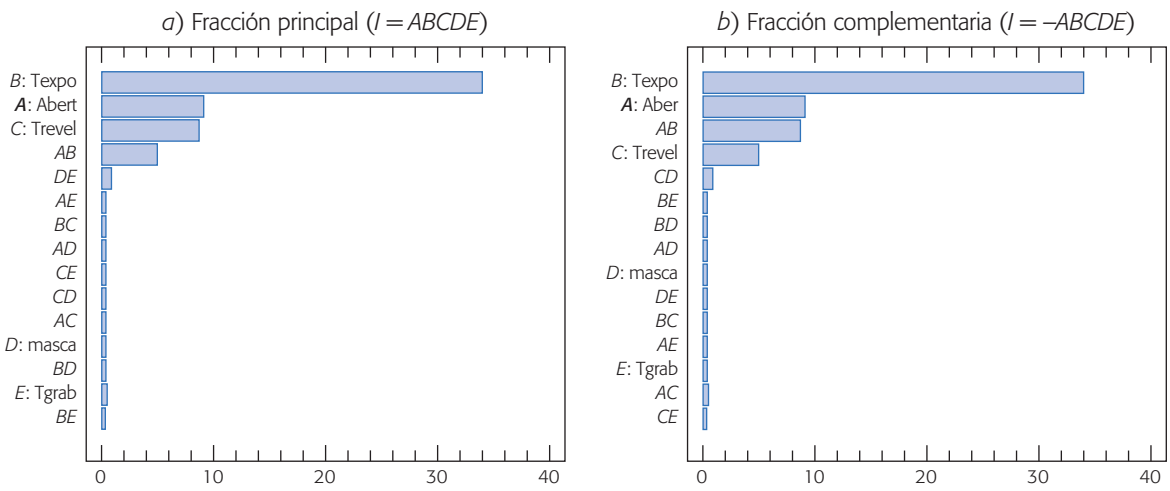


Figura 8.2 Diagramas de Pareto de efectos para las dos fracciones del experimento 2^5, ejemplo de semiconductores.

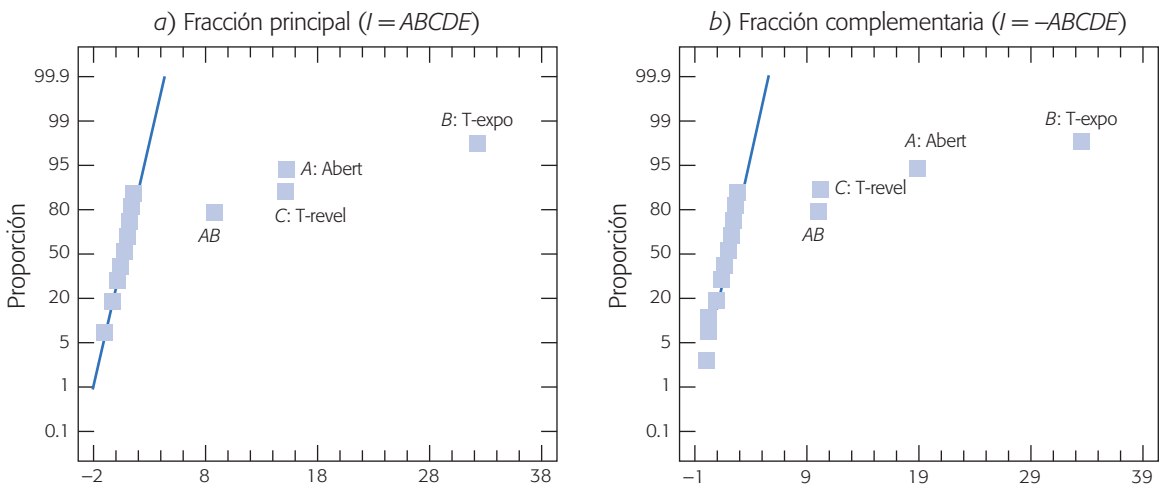


Figura 8.3 Gráficos de Daniel para las dos fracciones del diseño 2^5; ejemplo de semiconductores.

Interpretación

En la figura 8.4 se muestran los efectos principales detectados en cada una de las fracciones, y en la figura 8.5 se presenta el efecto de interacción *AB* de la manera que fue detectado por cada una de las fracciones. Las diferencias entre lo que se detecta con una fracción y otra es prácticamente imperceptible. Si comparamos estas gráficas contra las correspondientes del diseño factorial completo, notamos que desde el punto de vista físico se detecta que pasa lo mismo: a medida que se incrementan los niveles de los factores *A*, *B* y *C*, el rendimiento aumenta. Además, el efecto del nivel de abertura (*A*) es mayor cuando el tiempo de exposición es alto (*B*). Por lo tanto, al igual que con el factorial completo, el mejor tratamiento es (A^+, B^+, C^+, $D^\$$, $E^\$$). Los niveles de los factores *D* y *E* se eligen con el uso del criterio económico.

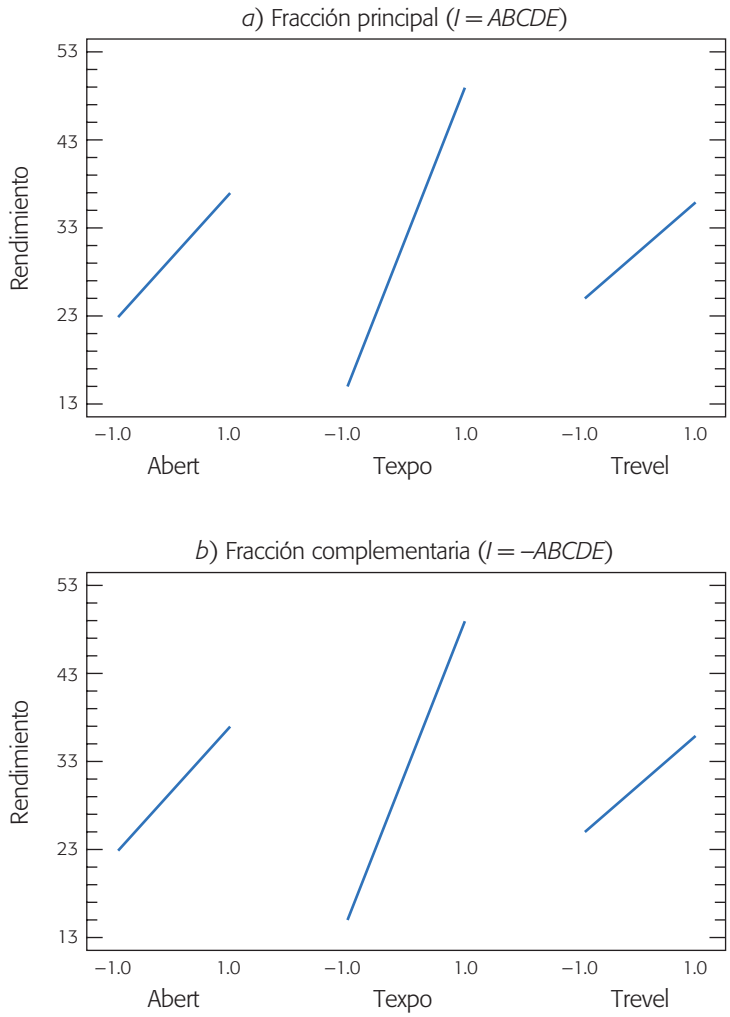


Figura 8.4 Efectos principales para las dos fracciones del diseño 2^5; ejemplo de semiconductores.

En resumen, el diseño factorial 2^5 sobre rendimiento de semiconductores se puede analizar de distintas maneras y los resultados son equivalentes. Se ha analizado como un $2^5(n = 1)$, como un 2^3 $(n = 4)$ usando la propiedad de colapsación, como una fracción 2_V^{5-1} $(I = ABCDE)$ y su complementaria 2_V^{5-1} $(I = -ABCDE)$. Algo importante a resaltar es que bastaba con una fracción a la mitad, y por lo tanto, la mitad de los recursos invertidos para obtener las mismas conclusiones. Se puede mostrar (véase ejercicio 6) que incluso con una fracción a la cuarta parte del diseño original es posible llegar a las mismas conclusiones con este ejemplo.

Diseños factoriales fraccionados 2^{k-2}

De acuerdo con el número de factores y el costo de cada corrida experimental, en ocasiones es necesario correr una fracción más pequeña del diseño factorial completo

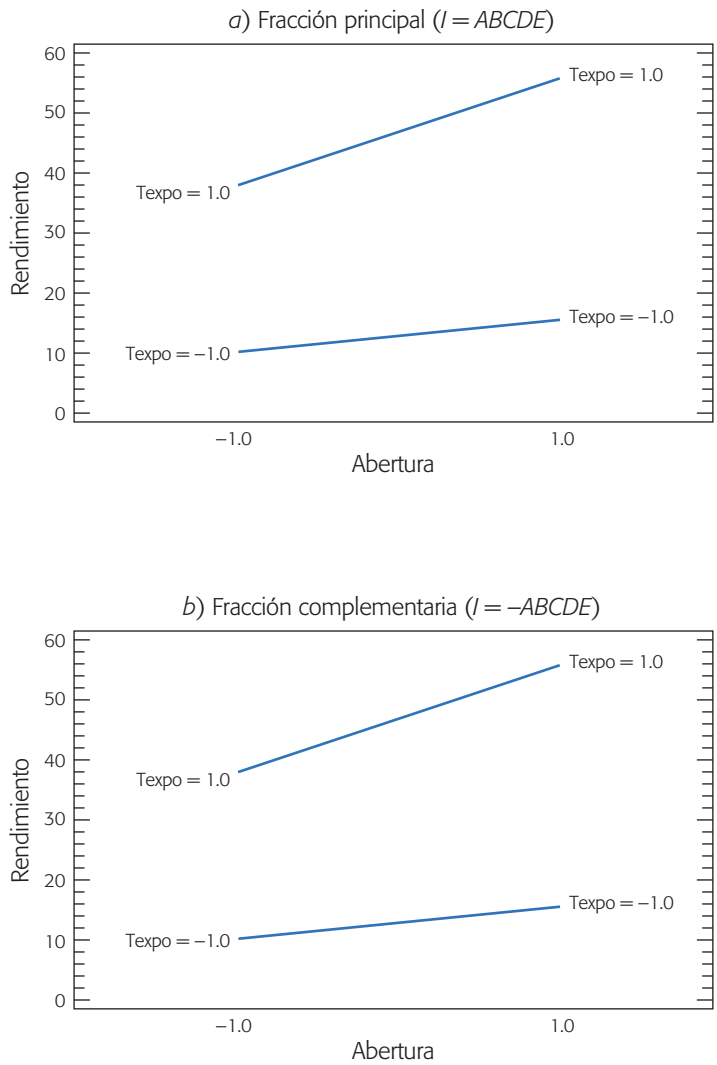


Figura 8.5 Interacción *AB* para las dos fracciones del diseño 2^5; ejemplo de semiconductores.

Efectos generadores
Son interacciones del más alto orden posible que son utilizadas para generar la fracción factorial. Su producto también debe ser una interacción del más alto orden posible.

2^k. Un diseño factorial fraccionado 2^{k-2} representa la cuarta parte del factorial original completo ($\frac{1}{4}\ 2^k = 2^{k-2}$). Para obtener este diseño se necesitan dos *efectos generadores*, de entre las interacciones del más alto orden. Estos efectos generadores deben ser elegidos de manera que su producto también sea una interacción del más alto orden posible. Estos diseños tendrán tres generadores: los primeros dos que se seleccionaron más su producto entre sí, y ninguno será estimable. La estructura alias se obtiene con estos tres generadores, de aquí que cada efecto tiene tres alias. En general, el número de "palabras" de la relación definidora indica la cantidad de alias que tendrá cada efecto, y multiplicando un efecto dado por esta relación se determinan sus alias. La palabra con menos letras en la relación definidora indica la resolución de la fracción.

El diseño factorial fraccionado 2^{k-2} se puede construir en dos pasos:

1. Se escribe el diseño 2^{k-2} como si fuera un factorial completo en $k-2$ factores, y de esta forma se tienen los niveles de los primeros $k-2$ factores.
2. Los niveles que corresponden a los factores de las dos últimas columnas (factores $k-1$ y k) se obtienen multiplicando columnas previas de acuerdo a los generadores.

Ejemplo 8.2

Fracción 2^{5-2}. Supongamos que se tienen cinco factores y que sólo se cuenta con recursos para correr una cuarta parte del diseño factorial 2^5, es decir, se quiere correr un diseño fraccionado 2^{5-2} que se construye mediante los dos pasos siguientes:

1. Se escribe el diseño factorial completo 2^3 para los tres primeros factores A, B y C, dejando las columnas D y E en blanco:

A	B	C	D	E
−	−	−		
+	−	−		
−	+	−		
+	+	−		
−	−	+		
+	−	+		
−	+	+		
+	+	+		

2. Los niveles para los factores D y E se obtienen al seleccionar de manera adecuada generadores. En este caso se proponen como los generadores $I = ABD$ e $I = ACE$, y el tercer generador es el producto $ABD \times ACE = BCDE$. Así, la relación definidora del diseño queda como $I = ABD = ACE = BCDE$. Al reescribir los generadores en la forma: $I = ABD \Rightarrow D = AB$ e $I = ACE \Rightarrow E = AC$, entonces con los productos AB y AC se generan los niveles del factor D y E, respectivamente. Haciendo los productos indicados se completa la tabla anterior y se tiene el diseño factorial que se muestra en la tabla 8.5, que tiene resolución III puesto que en la relación definidora el generador más pequeño tiene tres letras.

Tabla 8.5 Diseño 2^{5-2}, $I = ABD = ACE$.

A	B	C	D	E
−	−	−	+	+
+	−	−	−	−
−	+	−	−	+
+	+	−	+	−
−	−	+	+	−
+	−	+	−	+
−	+	+	−	−
+	+	+	+	+

Tabla 8.6 Estructura alias completa del diseño 2^{5-2}_{III}.

$A + BD + CE + ABCDE$
$B + AD + ABCE + CDE$
$C + ABCD + AE + BDE$
$D + AB + ACDE + BCE$
$E + ABDE + AC + BCD$
$BC + ACD + ABE + DE$
$BE + ADE + ABC + CD$
$\mu + ABD + ACE + BCDE$

Estructura de alias. La estructura de alias se obtiene al multiplicar cada efecto por la relación definidora dada por:

$$I = ABD = ACE = BCDE$$

y se muestra en la tabla 8.6. Los grupos de efectos alias ahora tienen cuatro elementos. Cada efecto principal cuenta con al menos una interacción doble como su alias, que es lo que implica la resolución III. No es posible construir una fracción 2^{5-2} con resolución mayor que III.

La estructura de alias completa incorpora hasta los efectos de interacción del más alto orden, aunque no sean de interés. Toda la información importante está contenida en la estructura de alias reducida, que involucra sólo hasta las interacciones dobles (véase tabla 8.7). Cuando se alían efectos con la misma jerarquía, como es el caso de $BE + CE$, debe decidirse con base en el conocimiento del proceso, a cuál interacción se atribuirá el efecto observado en caso de que resulte significativo. Otro criterio es fijarse cuáles efectos principales resultaron significativos, ya que éstos tienen más probabilidad de estar activos también en sus interacciones.

Con los generadores $I = ABD$ e $I = ACE$ se obtiene el diseño 2^{5-2} llamado fracción principal, que hemos presentado, pero tomando las otras combinaciones de signos en los generadores (($I = -ABD$, $I = ACE$), ($I = ABD$, $I = -ACE$) o ($I = -ABD$, $I = -ACE$)), se obtienen otros tres posibles diseños 2^{5-2} que pertenecen a la misma familia, y todos con la misma resolución. La decisión por una fracción en particular se puede hacer considerando los tratamientos que incluye cada fracción y viendo si alguno de éstos se prefiere o se quiere evitar, como por ejemplo, el tratamiento (+, +, +, +, +,).

Tabla 8.7 Estructura de alias reducida.

$A + BD + CE$
$B + AD$
$C + AE$
$D + AB$
$E + AC$
$BC + DE$
$BE + CE$

Diseño factorial fraccionado 2^{k-p}

Diseño factorial fraccionado 2^{k-p}
Es una fracción $1/2^p$ del factorial completo 2^k. Se construye con base en p generadores iniciales.

En general, un *diseño factorial fraccionado* 2^{k-p} es una fracción $1/2^p$ del diseño factorial completo 2^k. Para construir un diseño 2^{k-p} se eligen p generadores iniciales, todos interacciones del más alto orden posible, de manera que todos sus productos también sean interacciones de alto orden. Una vez elegidos los p generadores el diseño se puede construir en dos pasos, a saber: 1) Se escribe el diseño 2^{k-p} como si fuera el factorial completo para $k - p$ factores. 2) Para los últimos p factores las columnas de signos se obtienen multiplicando las columnas que indican los generadores.

La *relación definidora* tiene tantos términos como productos se puedan hacer con los p generadores. De la relación definidora se obtiene la estructura de alias y la resolución de la fracción resultante.

En todo el procedimiento descrito quizá lo más difícil es encontrar los mejores generadores de la fracción que se desea utilizar. Por fortuna, existen tablas de diseños factoriales fraccionados que incluyen su estructura de alias, lo que permite al investigador elegir el diseño, y saber con anticipación cuáles serían los alias de los efectos potencialmente importantes. Esta selección adecuada del diseño factorial fraccionado también se puede hacer con software estadísticos.

En la tabla 8.8 se proporcionan los generadores de algunos diseños factoriales fraccionados ($k \leq 10$), que al menos tienen resolución IV. También se da el grado de fraccionamiento, el número de alias que tiene cada efecto, el número de tratamientos y el número de factores que se quieren estudiar. Los diseños de resolución III no incluidos en la tabla, también son sumamente útiles, en particular cuando se tienen más de ocho factores y resulta costoso correr el diseño de resolución IV.

Ejemplo 8.3

Fracción 2^{7-3}. Supongamos que se quieren estudiar $k = 7$ factores y sólo se tienen recursos para correr una octava parte del diseño 2^7 completo, por lo que se decide utilizar un diseño factorial fraccionado 2^{7-3}. En la tabla 8.8 se lee que los generadores $E = \pm ABC$, $F = \pm BCD$ y $G = \pm ACD$ son adecuados. Se pueden construir ocho fracciones diferentes con estos generadores, dependiendo de los signos que se toman. Por facilidad vamos a construir la fracción principal, que es la que se obtiene con los tres generadores con signo positivo.

En un primer paso se escribe el diseño $2^{7-3} = 2^4$ como si fuera un factorial completo: ver las primeras cuatro columnas de la tabla 8.9. En el segundo paso se calculan las columnas faltantes E, F y G con base en los generadores. El diseño de la tabla 8.9 es el factorial fraccionado 2^{7-3} deseado.

La estructura de alias completa del diseño se obtiene multiplicando cada efecto por la relación definidora, que está dada por los generadores iniciales y todos sus posibles productos: $I = ABCE = ABFG = ACDG = ADEF = BCDF = BDEG = CEFG$. Así, la estructura de alias completa se muestra en la tabla 8.10, donde cada efecto tiene siete alias. Ésta tiene estructura de poca o nula utilidad, salvo tener la idea de la situación que se presenta con los efectos. Para fines prácticos basta conocer la estructura de alias reducida que se muestra en la tabla 8.11.

De acuerdo con esta tabla, en este diseño habría que cuidar que dos interacciones de interés especial para el experimentador no aparezcan alias entre ellas. Esto se

Tabla 8.8 Factoriales fraccionados con resolución *IV*, con máximo 64 corridas.

Número de factores	Diseño	Número de corridas	Alias de cada efecto	Grado de fracción	Generadores
4	2_{IV}^{4-1}	8	1	1/2	$D = \pm ABC$
5	2_{V}^{5-1}	16	1	1/2	$E = \pm ABCD$
6	2_{VI}^{6-1}	32	1	1/2	$F = \pm ABCDE$
6	2_{IV}^{6-2}	16	3	1/4	$E = \pm ABC$ $F = \pm BCD$
7	2_{VII}^{7-1}	64	1	1/2	$G = \pm ABCDEF$
7	2_{IV}^{7-2}	32	3	1/4	$F = \pm ABCD$ $G = \pm ABDE$
7	2_{IV}^{7-3}	16	7	1/8	$E = \pm ABC$; $F = \pm BCD$ $G = \pm ACD$
8	2_{V}^{8-2}	64	3	1/4	$G = \pm ABCD$ $H = \pm ABEF$
8	2_{IV}^{8-3}	32	7	1/8	$F = \pm ABC$; $G = \pm ABD$ $H = \pm BCDE$
8	2_{IV}^{8-4}	16	15	1/16	$E = \pm BCD$; $F = \pm ACD$ $G = \pm ABC$; $H = \pm ABD$
9	2_{IV}^{9-3}	64	7	1/8	$G = \pm ABCD$; $\pm H = \pm ACEF$ $J = \pm CDEF$
9	2_{IV}^{9-4}	32	15	1/16	$F = \pm BCDE$; $G = \pm ACDE$ $H = \pm ABDE$; $J = \pm ABCE$
10	2_{IV}^{10-4}	64	15	1/16	$G = \pm BCDF$; $H = \pm ACDF$ $J = ABDE$; $K = \pm ABCE$
10	2_{IV}^{10-5}	32	31	1/32	$F = \pm ABCD$; $G = \pm ABCE$; $H = \pm ABDE$ $J = \pm ACDE$; $K = \pm BCDE$

Tabla 8.9 Diseño 2^{7-3}, con generadores $I = ABCE$, $I = BCDF$ e $I = ACDG$.

A	*B*	*C*	*D*	*E* = *ABC*	*F* = *ADE*	*G* = *ACD*
−	−	−	−	−	−	−
+	−	−	−	+	−	+
−	+	−	−	+	+	−
+	+	−	−	−	+	+
−	−	+	−	+	+	+
+	−	+	−	−	+	−
−	+	+	−	−	−	+
+	+	+	−	+	−	−
−	−	−	+	−	+	+
+	−	−	+	+	+	−
−	+	−	+	+	−	+
+	+	−	+	−	−	−
−	−	+	+	+	−	−
+	−	+	+	−	−	+
−	+	+	+	−	+	−
+	+	+	+	+	+	+

logra al asignar los factores a las letras de las columnas (tabla 8.9), de manera que las interacciones de interés aparezcan en grupos de alias diferentes. Por ejemplo, supongamos que cuatro de los siete factores son: *temperatura*, *humedad*, *velocidad* y

Tabla 8.10 Estructura de alias completa del diseño 2_{IV}^{7-3}.

$$A + BCE + BFG + CDG + DEF + ABCDF + ABDEG + ACEFG$$
$$B + ACE + AFG + CDF + DEF + ABCDG + ABDEF + BCEFG$$
$$C + ABE + ADG + BDF + EFG + ABCFG + ACDEF + BCDEG$$
$$D + ACG + AEF + BCF + BEG + ABCDE + ABDFG + CDEFG$$
$$E + ABC + ADF + BDG + CFG + ABEFG + ACDEG + BCDEF$$
$$F + ABG + ADE + BCD + CEG + ABCEF + ACDFG + BDEFG$$
$$G + ABF + ACD + BDE + CEF + ABCEG + ADEFG + BCDFG$$
$$AB + CE + FG + ACDF + ADEG + BCDG + BDEF + ABCEFG$$
$$AC + BE + DG + ABDF + AEFG + BCFG + CDEF + ABCDEG$$
$$AD + CG + EF + ABCF + ABEG + BCDE + BDFG + ACDEFG$$
$$AE + BC + DF + ABDG + ACFG + BEFG + CDEG + ABCDEF$$
$$AF + BG + DE + ABCD + ACEG + BCEF + CDFG + ABDEFG$$
$$AG + FG + CD + ABDE + ACEF + BCEG + DEFG + ABCDFG$$
$$BD + CF + EG + ABCG + ABEF + ACDE + ADFG + BCDEFG$$
$$ABD + ACF + AEG + BCG + BEF + CDE + DFG + ABCDEFG$$
$$\mu + ABCE + ABFG + ACDG + ADEF + BCEF + BDEF + CEFG$$

Tabla 8.11 Estructura de alias reducida para el diseño 2_{IV}^{7-3}.

Los efectos principales: A, B, C, D, E, F y G se estiman limpiamente, pero las interacciones dobles se confunden entre ellas:

$$AB + CE + FG$$
$$AC + BE + DG$$
$$AD + CG + EF$$
$$AE + BC + DF$$
$$AF + BG + DE$$
$$AG + BF + CD$$
$$BD + CF + EG$$

ángulo, e interesa estimar todas las interacciones dobles entre estos factores; además, supongamos que de los otros tres factores sólo interesan los efectos principales. Se asigna la letra A al factor temperatura, B al factor humedad, C a velocidad y D a ángulo, y con esto se podrán estimar las seis interacciones de interés. En contraparte, si los primeros tres factores se asignan de igual manera pero el factor ángulo se asigna a la columna E, se tendrán varias interacciones de interés alias entre ellas, como son: $AB + CE$, $AC + BE$ y $AE + BC$. Hay que observar que un pequeño cambio en la asignación de los factores a las columnas del diseño, cambia radicalmente la estructura de alias en cuanto a los efectos de interés.

Note que cuando todas las interacciones de dos factores son igual de importantes, este diseño no hace un buen trabajo y debe utilizarse otro con resolución al menos de V. Pero si de antemano se tienen identificadas sólo algunas de las interacciones como las potencialmente importantes, se puede buscar la manera de estudiarlas con una fracción de resolución IV.

Estimación de efectos y sumas de cuadrados. Los efectos y las sumas de cuadrados en los diseños factoriales fraccionados 2^{k-p} se obtienen a partir de los contrastes, de manera similar a como se hace con los factoriales completos 2^k. Se obtiene un

contraste para cada grupo de efectos alias y se pondera por una constante apropiada para estimar el efecto correspondiente como una diferencia de medias. Así, el efecto de un grupo de efectos alias *X* se estima como:

$$Efecto\ X = \frac{(Contraste X)}{2^{k-p-1}}$$

y su correspondiente suma de cuadrados es:

$$SC_X = \frac{(Contraste X)^2}{2^{k-p}}$$

la cual tiene un grado de libertad. Al calcular la suma de cuadrados totales (SC_T) y la suma de cuadrados del error (SC_E), compuesta por efectos pequeños excluidos con base en el Pareto efectos y el gráfico de Daniel, se puede construir un ANOVA razonable. Cabe señalar que para excluir un efecto del análisis se deben excluir también todos sus alias.

Experimento 2^{7-4}: ejemplo integrador

En una compañía se compró un equipo de afilado, y después de una semana se dieron cuenta de que el exceso de vibración en el proceso era un problema muy serio. Un equipo de mejora decide tratar de reducir la vibración aplicando diseño de experimentos. Identifican *siete factores*, todos asociados con la herramienta de afilado, que pueden tener algo que ver en la cantidad de vibración: *diámetro*, *longitud*, *tamaño de grano del material*, *velocidad de alimentación*, *revoluciones por minuto*, *estructura del material* y *peso de la precarga*. Se eligen dos niveles para cada factor, en los cuales llevar a cabo el experimento. Como la puesta en marcha del robot y el tiempo de corrida tienen un alto costo, el equipo decide utilizar un experimento de ocho corridas. Tanto los factores como los niveles utilizados en unidades originales se muestran en la tabla 8.12.

Bajo el supuesto de que los efectos de interacción se pueden considerar despreciables, se decide correr un diseño altamente fraccionado (saturado) como lo es la fracción 2^{7-4}_{III} que se muestra en la tabla 8.14 junto con los resultados obtenidos. El diseño tiene *resolución III*, lo cual implica que hay efectos principales confundidos con interacciones dobles. La estructura de alias reducida se muestra en la tabla 8.13.

Análisis del experimento

El experimento tiene un total de siete grados de libertad, que se gastan en estimar sólo a los efectos principales, y quedan cero grados de libertad para el error en el ANOVA. De aquí la necesidad de recurrir al diagrama de Pareto y al gráfico de Daniel como paso previo antes de intentar un análisis de varianza.

El diagrama de Pareto de efectos hace un trabajo excelente al detectar tres efectos importantes y cuatro despreciables (véase figura 8.6). Esto se ve confirmado por el gráfico de Daniel de la figura 8.7, que muestra cuatro efectos sobre la línea (que no afectan) y tres efectos separados de la línea, que corresponden a las tres barras más altas en el Pareto.

Tabla 8.12 Factores y niveles utilizados: problema de vibración.

Factor	Descripción (unidades)	Niveles (bajo, alto)
C: diam	Diámetro (pulgadas)	1.0, 1.5
B: long	Longitud (pulgadas)	1.0, 2.0
A: grano	Tamaño de grano (/ pulgada)	80, 120
G: alim	Velocidad de alimentación (pulg/min)	2.0, 4.0
D: rpm	Rpm (× 1 000)	15, 20
E: precar	Peso de precarga (libras)	1.0, 0.4
F: matest	Estructura del material (onzas)	1.0, 4.0

Tabla 8.13 Estructura de alias del diseño 2^{7-4}_{III} (fracción principal).

$A + BD + CE + FG$
$B + AD + CF + EG$
$C + AE + BF + DG$
$D + AB + CG + EF$
$E + AC + BG + DF$
$F + AG + BC + DE$
$G + AF + BE + CD$

Tabla 8.14 Matriz de diseño y vibración observada.

Grano	Long.	Diám.	RPM	Precar	Matest	Alim.	Vibra
–1.0	–1.0	–1.0	1.0	1.0	1.0	–1.0	77.4
1.0	–1.0	–1.0	–1.0	–1.0	1.0	1.0	68.3
–1.0	1.0	–1.0	–1.0	1.0	–1.0	1.0	81.9
1.0	1.0	–1.0	1.0	–1.0	–1.0	–1.0	66.2
–1.0	–1.0	1.0	1.0	–1.0	–1.0	1.0	42.1
1.0	–1.0	1.0	–1.0	1.0	–1.0	–1.0	78.3
–1.0	1.0	1.0	–1.0	–1.0	1.0	–1.0	39.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	68.4

Con esta información se puede intentar un análisis de varianza, mandando al error a los cuatro efectos alineados. El resultado se muestra en la tabla 8.15.

El estadístico R^2_{aj} que mide el grado de explicación de la variable de respuesta por el modelo es 93.71%. Aunque la suma de cuadrados del error tiene sólo cuatro grados de libertad, en este caso, parece que el cuadrado medio del error está estimado de manera correcta y razonable.

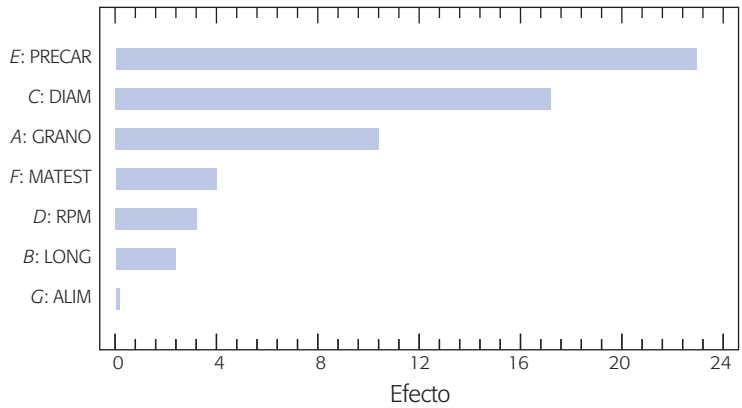


Figura 8.6 Pareto de efectos para el problema de vibración.

Tabla 8.15 Mejor análisis de varianza.

FV	SC	GL	CM	F_0	Valor-*p*
A: grano	208.08	1	208.08	12.58	0.0239
C: diám.	544.5	1	544.5	32.93	0.0046
E: precar	1 021.52	1	1 021.52	61.78	0.0014
Error	66.14	4	16.535		
Total	1 840.24	7			

Interpretación

El objetivo del estudio es encontrar una combinación de los niveles de los factores que minimicen el problema de la vibración de la máquina de afilado. En la figura 8.8 se muestran los efectos que están activos, y para cada uno se localiza el valor más bajo de la línea. De esta forma se aprecia que a mayor tamaño de precarga mayor vibración, a más diámetro menor vibración y a mayor tamaño de grano más vibra-

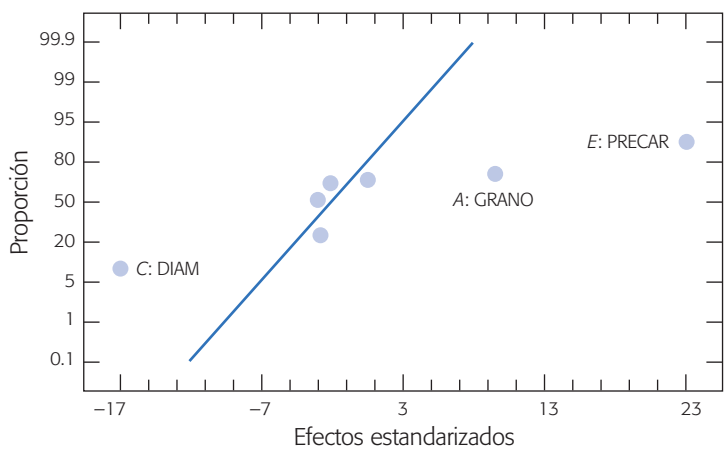


Figura 8.7 Efectos en papel normal para el problema de vibración.

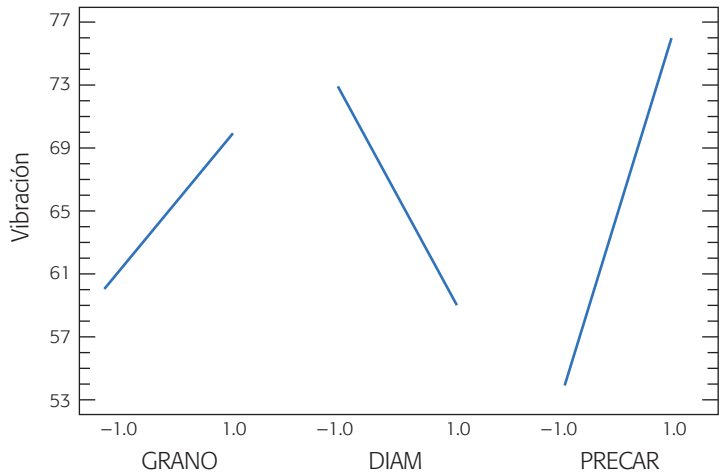


Figura 8.8 Representación de efectos activos para el problema de vibración.

ción. Por lo tanto, el mejor tratamiento es: GRANO en su nivel bajo, DIAM en su nivel alto y PRECAR en su nivel bajo; los niveles de los factores restantes se eligen con el criterio de economía o productividad.

Como sólo hay tres factores activos tiene sentido dibujar la gráfica de cubo, que resume bien el comportamiento de la vibración predicha (véase figura 8.9). Se observa que la vibración mínima (40.55) ocurre en la combinación *(A*: *GRANO*$^{-}$, *C*: *DIAM*$^{+}$, *E*: *PRECAR*$^{-}$).

Tópicos adicionales sobre factoriales fraccionados

Comentarios sobre la resolución

En general, no es recomendable correr diseños de resolución III, a menos que se esté dispuesto a aceptar que sólo importan los efectos principales. En algunos procesos es arriesgado suponer de antemano que ninguna interacción doble está activa. Sin embargo, con frecuencia surgen situaciones en las que se debe utilizar un diseño de resolución III. Por ejemplo, cuando se tiene una gran cantidad de factores a estudiar no queda otra alternativa si se quiere un número razonable de corridas experimentales. Aun con pocos factores, cuando cada corrida del proceso es demasiado cara, es necesario recurrir a una fracción de resolución III.

En la práctica se busca tener la máxima resolución posible, con un número razonable de corridas experimentales y de gasto de recursos. Con una cantidad no muy grande de factores ($5 < k \leq 15$) existen fracciones de resolución IV que no requieren más de 32 corridas experimentales, y proporcionan información de todos los efectos principales y de algunas de las interacciones dobles.

Si el proceso es masivo y las corridas del mismo son baratas, se puede optar directamente por fracciones de resolución *V*; es en este tipo de procesos donde es admisible que se incremente el número de corridas experimentales hasta 64 o más.

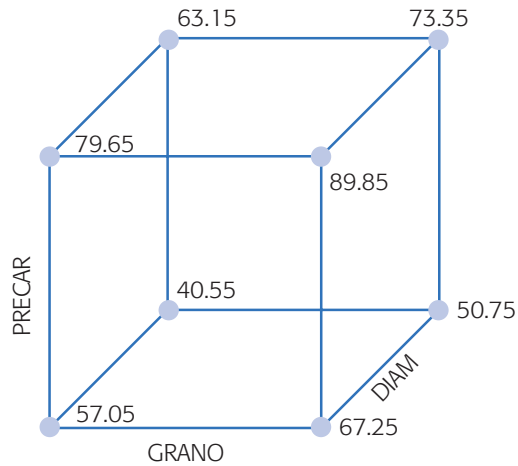


Figura 8.9 Gráfica de cubo y vibración predicha en cada tratamiento.

Diseños fraccionados con aberración mínima

El que dos diseños factoriales fraccionados tengan la misma resolución no significa que posean la misma habilidad para estimar los efectos potencialmente importantes. Por ejemplo, consideremos los dos diseños 2^{7-2}, cuyas relaciones definidoras son:

$$\begin{aligned} d_1&: I = DEFG = ABCDF = ABCEG \\ d_2&: I = ABCF = ADEG = BCDEFG \end{aligned} \qquad \textbf{(8.3)}$$

Ambos diseños son de resolución IV, sin embargo, mientras que el diseño d_1 tiene tres pares de interacciones dobles que son alias, el diseño d_2 tiene seis pares. Así, aunque ambos diseños poseen la misma resolución, el primero permite estimar limpiamente más interacciones de dos factores.

Patrón de longitud de palabra. El patrón de longitud de palabra para un diseño factorial fraccionado 2^{k-p} está dado por $W = (A_1, A_2, \ldots, A_k)$, donde A_i es el número de palabras con i letras en la relación definidora. Por ejemplo, para el diseño d_1 el patrón de longitud de palabra es $W(d_1) = (0, 0, 0, 1, 2, 0, 0)$, ya que no hay ninguna relación definidora con 1, 2 y 3 letras, hay una con cuatro letras y dos con cinco.

Aberración mínima. Se dice que d_{MA} es la fracción con aberración mínima si $A_r(d_{MA}) < A_r(d_i)$ para toda i, donde r es el número de letras de la palabra más pequeña en la relación definidora de cada diseño.

Por ejemplo, los dos diseños 2^{7-2} definidos por las relaciones definidoras de la ecuación 8.3, tienen patrones de longitud de palabra $W(d_1) = (0, 0, 0, 1, 2, 0, 0)$ y $W(d_2) = (0, 0, 0, 2, 0, 1, 0)$, respectivamente. Por lo tanto, el primero tiene menor aberración que el segundo, ya que $A_4(d_1) = 1 < A_4(d_2) = 2$. Es decir, el diseño d_1 tiene una relación definidora con cuatro letras, mientras que el diseño d_2 tiene dos. Así, el concepto de *mínima aberración* es un criterio que casi siempre permite escoger el mejor diseño 2^{k-p} entre las posibles fracciones con la misma resolución.

Mínima aberración
Criterio que permite escoger un diseño 2^{k-p} entre las posibles fracciones con la misma resolución.

Fracciones saturadas. Un *diseño factorial fraccionado* se llama *saturado* cuando el número total de grados de libertad del experimento es igual al número de factores que se estudian. Esto es, todos los grados de libertad disponibles se gastan en los efectos principales. Por lo tanto, estos diseños tienen resolución III, ya que los efectos principales se confunden con las interacciones dobles.

Factorial fraccionado saturado
Diseño que estudia los efectos principales de k factores usando $k + 1$ corridas experimentales.

Un ejemplo de diseño factorial fraccionado saturado lo constituye el diseño 2^{7-4}_{III} que se utilizó en el ejemplo de la vibración del proceso de afilado (véase tabla 8.14).

Con este diseño sólo se pueden estudiar los efectos principales, suponiendo *a priori* que las interacciones dobles no son importantes. Esto conlleva un riesgo que se puede correr cuando de plano no se tienen los recursos necesarios para hacer más de ocho pruebas. En general es más recomendable experimentar aunque sea con resolución III que no hacerlo.

Diseños de Plackett-Burman

Diseños de Plackett-Burman
Este diseño representa otra alternativa para fraccionar factoriales completos 2^k, donde el número de puntos de diseño no necesariamente es potencia de dos, pero sí múltiplo de cuatro.

Los *diseños de Plackett-Burman* representan otra alternativa para fraccionar diseños factoriales completos 2^k, donde el número de puntos de diseño no necesariamente es potencia de dos pero sí es múltiplo de cuatro. En un momento dado estas nuevas fracciones permiten optimizar los recursos disponibles. Los diseños de Plackett-Burman son fracciones del diseño factorial 2^k, donde el número de puntos de diseño N es múltiplo de cuatro. Cuando N es potencia de 2, estos diseños son idénticos a los fraccionados 2^{k-p} antes descritos.

Para construir un diseño de Plackett-Burman para k factores se selecciona un renglón o columna de niveles codificados –1 y 1, de manera que el número de positivos sea $(k + 1)/2$ y el de negativos $(k – 1)/2$. Este renglón es el primero del diseño. Los siguientes $k – 1$ renglones o columnas se generan recorriendo cada vez un lugar el primer renglón (ver tabla 8.16). Por último, el $(k + 1)$-ésimo renglón se forma sólo de números –1 (todos los factores en su nivel bajo).

Diseño de Plackett-Burman con 12 corridas y hasta $k = 11$ factores. Si el primer renglón es (+1, –1, +1, –1, –1, –1, +1, +1, +1, –1, +1) el diseño de Plackett-Burman generado a partir de éste se muestra en la tabla 8.16. Observe que los renglones del 2 al 11 resultan al desplazar a la derecha el primer renglón. Finalmente, en la combinación 12 todos los factores se fijan en su nivel bajo.

Cabe aclarar que no cualquier orden en los signos +, – del primer renglón da por resultado un diseño adecuado (véase Box, Hunter y Hunter, 1988). En la tabla 8.17 se muestran los "primeros renglones" (o columnas) que se proponen para los diseños de Plackett-Burman más utilizados. En ella, k es el número máximo de factores y N es el número de tratamientos de diseño. Aunque el arreglo es fácil de construir a partir de rotar los signos propuestos, resulta más práctico generarlo con un software.

Si el número de factores a estudiar es menor que k, una vez construido el diseño de la forma que se explicó se seleccionan tantas columnas como factores. Por ejemplo, si con el diseño de la tabla 8.16 se quieren estudiar sólo seis factores, entonces la matriz de diseño estará formada por las primeras seis columnas.

Tabla 8.16 Diseño de Plackett-Burman con 12 corridas y hasta $k = 11$ factores.

A	B	C	D	E	F	G	H	I	J	K
1	–1	1	–1	–1	–1	1	1	1	–1	1
1	1	–1	1	–1	–1	–1	1	1	1	–1
–1	1	1	–1	1	–1	–1	–1	1	1	1
1	–1	1	1	–1	1	–1	–1	–1	1	1
1	1	–1	1	1	–1	1	–1	–1	–1	1
1	1	1	–1	1	1	–1	1	–1	–1	–1
–1	1	1	1	–1	1	1	–1	1	–1	–1
–1	–1	1	1	1	–1	1	1	–1	1	–1
–1	–1	–1	1	1	1	–1	1	1	–1	1
1	–1	–1	–1	1	1	1	–1	1	1	–1
–1	1	–1	–1	–1	1	1	1	–1	1	1
–1	–1	–1	–1	–1	–1	–1	–1	–1	–1	–1

Aclaración de ambigüedades con otra fracción

En ocasiones los resultados de un diseño factorial altamente fraccionado son ambiguos, en el sentido de que al momento del análisis surgen dudas sobre la interpretación de efectos que son alias, por lo que se hace necesario correr otra fracción que aclare las ambigüedades, separando de los grupos de alias aquellos efectos que son de interés para el experimentador. La idea es combinar de manera adecuada la información de ambas fracciones para separar los efectos deseados.

Estas ambigüedades pueden ser el resultado de una mala planeación del experimento. Por ejemplo, cuando el grupo de trabajo aun a sabiendas de que la interacción BC tiene muchas posibilidades de afectar la variable de interés, decide correr un diseño de resolución III en el que BC se confundirá con el efecto A. Aquí, la necesidad de correr una fracción adicional se pudo evitar desde el principio, seleccionando un diseño con menor grado de fraccionamiento que permitiera estimar por separado el efecto BC. Sin embargo, lo típico es que la ambigüedad surja después de analizar la primera fracción y observar que los resultados obtenidos son interesantes; no obstante, se plantea la necesidad de correr una fracción adicional para tener una visión más clara de lo que sucede con los efectos de interés.

La fracción adicional debe elegirse cuidando que en realidad sirva para resolver las imprecisiones heredadas de la primera fracción. El truco es elegir otra fracción que, al combinarse ambas estructuras alias, aclare las dudas. Por ejemplo, si en la primera fracción se confunde A con BC de la forma $A + BC$, y se desea separarlos

Tabla 8.17 Signos para el primer renglón de algunos diseños de Plackett-Burman.

$k = 11, N = 12$	+ – + – – – + + + – +
$k = 19, N = 20$	+ – + + – – – – + – + – + + + + – – +
$k = 23, N = 24$	+ – – – – + – + – – + + – – + + – + – + – + + + +
$k = 35, N = 36$	– – + – – + + – + – + – – – – + – – + + + – + + + + + + – – – + + + – +

corriendo otra fracción, esta última deberá aliar los efectos en la forma $A - BC$, de manera que al combinar estos grupos (sumando y dividiendo entre dos), resulta:

$$\frac{(A + BC) + (A - BC)}{2} = A$$

y restando

$$\frac{(A + BC) - (A - BC)}{2} = BC$$

con lo que A y BC están limpiamente estimados. En resumen, dependiendo del tipo de confusión que se quiera eliminar, va a convenir restar o sumar las estructuras alias y luego dividir entre dos.

Son varias las situaciones de interés que podrían presentarse en la práctica. Aquí se presentan dos de ellas. Para ilustrar los dos casos se utiliza el diseño factorial fraccionado 2^{7-4} con generadores iniciales $I = \pm ABD = \pm ACE = \pm BCF = \pm ABCG$, el cual, en su fracción principal, se muestra en la tabla 8.18. La relación definidora de este diseño tiene 15 palabras que resultan de considerar los cuatro generadores iniciales y todos los posibles productos entre ellos. De aquí se desprende que cada efecto tiene 15 alias, o lo que es igual, los grupos de efectos alias tienen 16 miembros.

Este diseño, construido con los signos positivos en los cuatro generadores, es una de las 16 posibles fracciones que pueden construirse con las combinaciones de los signos de los cuatro generadores iniciales. Suponga que las interacciones de tres o más factores no son importantes, la estructura alias se muestra en la tabla 8.19. Denotemos con l_A a la suma estimada de los efectos alias que constituyen el grupo al que pertenece el efecto principal A, es decir, $l_A \rightarrow A + BD + CE + FG$, donde la flecha $\rightarrow$ significa "es estimador de". En la tabla 8.19 se muestra esta notación para cada grupo de efectos alias obtenidos con la fracción inicial.

CASO 1

Estimación de un factor dominante y todas sus interacciones. Suponga que al correr el diseño factorial fraccionado dado en la tabla 8.18 se observa, a la hora del análisis, que uno de los factores tiene un efecto fuerte y surge el interés de investigar si interactúa con los otros factores. Para ello se decide correr una fracción adicional

Tabla 8.18 Diseño factorial fraccionado 2^{7-4}.

A	B	C	D = AB	E = AC	F = BC	G = ABC	Total
−	−	−	+	+	+	−	*def*
+	−	−	−	−	+	+	*afg*
−	+	−	−	+	−	+	*beg*
+	+	−	+	−	−	−	*abd*
−	−	+	+	−	−	+	*cdg*
+	−	+	−	+	−	−	*ace*
−	+	+	−	−	+	−	*bcf*
+	+	+	+	+	+	+	*abcdefg*

Tabla 8.19 Estructura alias de la fracción original.

Estimador → Efecto alias
$l_A \rightarrow A + BD + CE + FG$
$l_B \rightarrow B + AD + CF + EG$
$l_C \rightarrow C + AE + BF + DG$
$l_D \rightarrow D + AB + CG + EF$
$l_E \rightarrow E + AC + BG + DF$
$l_F \rightarrow F + AG + BC + DE$
$l_G \rightarrow G + AF + BE + CD$

que permita estimar, al combinarse con los resultados de la primera, el efecto principal de ese factor dominante y todas las interacciones dobles en las que participa. La fracción adicional que se requiere se logra cambiando el signo a los generadores que involucran a este factor dominante. Por ejemplo, si el factor dominante es *A*, se utilizan los generadores dados por:

$$I = -ABD = -ACE = +BCF = -ABCG$$

La nueva fracción que se corre con estos generadores tiene la estructura de alias que se muestra en la tabla 8.20, que es la misma estructura alias de la primera fracción pero con algunos signos cambiados.

Al combinar ambas estructuras alias (tablas 8.19 y 8.20) se obtienen limpiamente todos los efectos que involucran al factor dominante *A*, pero los demás factores siguen confundidos con interacciones de dos factores. La nueva estructura de confusión se obtiene al sumar y restar las estructuras alias de las dos fracciones. El resultado se muestra en la tabla 8.21.

De esta manera se logra separar el efecto principal del factor dominante *A* y todas las interacciones en las que participa, pero el experimento total sigue siendo de resolución III.

CASO 2

Estimación de todos los efectos principales. Consideremos otra vez como primera fracción la de la tabla 8.18, pero ahora, dados los resultados del análisis surge la inquietud de estimar limpiamente los efectos principales, para estar más seguros de las conclusiones obtenidas.

Tabla 8.20 Estructura alias de la segunda fracción, caso 1.

Estimador → Efecto alias
$l_{A'} \rightarrow A - BD - CE - FG$
$l_{B'} \rightarrow B - AD + CF + EG$
$l_{C'} \rightarrow C - AE + BF + DG$
$l_{D'} \rightarrow D - AB + CG + EF$
$l_{E'} \rightarrow E - AC + BG + DF$
$l_{F'} \rightarrow F - AG + BC + DE$
$l_{G'} \rightarrow G - AF + BE + CD$

Tabla 8.21 Estructura alias combinada, caso 1.

Estimador → Efecto alias	Estimador → Efecto alias
$\frac{1}{2}(l_A - l_{A'}) \rightarrow BD + CE + FG$	$\frac{1}{2}(l_A + l_{A'}) \rightarrow A$
$\frac{1}{2}(l_B - l_{B'}) \rightarrow AD$	$\frac{1}{2}(l_B + l_{B'}) \rightarrow B + CF + EG$
$\frac{1}{2}(l_C - l_{C'}) \rightarrow AE$	$\frac{1}{2}(l_C + l_{C'}) \rightarrow C + BF + DG$
$\frac{1}{2}(l_D - l_{D'}) \rightarrow AB$	$\frac{1}{2}(l_D + l_{D'}) \rightarrow D + CG + EF$
$\frac{1}{2}(l_E - l_{E'}) \rightarrow AC$	$\frac{1}{2}(l_E + l_{E'}) \rightarrow E + BG + DF$
$\frac{1}{2}(l_F - l_{F'}) \rightarrow AG$	$\frac{1}{2}(l_F + l_{F'}) \rightarrow F + BC + DE$
$\frac{1}{2}(l_G - l_{G'}) \rightarrow AF$	$\frac{1}{2}(l_G + l_{G'}) \rightarrow G + BE + CD$

Recordemos que de acuerdo con la jerarquía de los efectos, al utilizar la fracción de la tabla 8.18, el experimentador estaría atribuyendo el efecto observado al efecto principal e ignorando las interacciones dobles. Sin embargo, siempre queda la inquietud de si el efecto observado en realidad se debe al efecto principal y no a alguna interacción doble del grupo de alias, y esto se puede aclarar corriendo una fracción adicional apropiada. La fracción adecuada en este caso es aquella que tiene por generadores los mismos de la primera fracción, pero con los signos de los generadores de tres letras cambiados. Esto es, la fracción adicional se genera con:

$$I = -ABD = -ACE = -BCF = +ABCG$$

que da lugar a la estructura alias de la tabla 8.22.

Si esta estructura alias se combina con la estructura alias de la primera fracción dada en la tabla 8.19, se obtiene la nueva estructura de confusión dada en la tabla 8.23. Se observa que los efectos principales están separados de las interacciones dobles, que era el objetivo que se perseguía. Note que las fracciones combinadas dan por resultado un diseño fraccionado de resolución IV, al no quedar ningún efecto principal confundido con alguna interacción doble.

Uso de software

Los diseños factoriales fraccionados en *Statgraphics* se construyen con la secuencia *Special* → *Experimental design* → *Create design* → *Screening* (o Doe → Design creation → Create new Design → Screening). Se declaran las variables de respuesta

Tabla 8.22 Estructura alias de la segunda fracción, caso 2.

Estimador → Efecto alias
$l_A \rightarrow A - BD - CE - FG$
$l_B \rightarrow B - AD - CF - EG$
$l_C \rightarrow C - AE - BF - DG$
$l_D \rightarrow D - AB - CG - EF$
$l_E \rightarrow E - AC - BG - DF$
$l_F \rightarrow F - AG - BC - DE$
$l_G \rightarrow G - AF - BE - CD$

Tabla 8.23 Estructura de alias combinada, caso 2.

Estimador → Efecto alias	Estimador → Efecto alias
$\frac{1}{2}(l_A + l_{A'}) \rightarrow A$	$\frac{1}{2}(l_A - l_{A'}) \rightarrow BD + CE + FG$
$\frac{1}{2}(l_B + l_{B'}) \rightarrow B$	$\frac{1}{2}(l_B - l_{B'}) \rightarrow AD + CF + EG$
$\frac{1}{2}(l_C + l_{C'}) \rightarrow C$	$\frac{1}{2}(l_C - l_{C'}) \rightarrow AE + BF + DG$
$\frac{1}{2}(l_D + l_{D'}) \rightarrow D$	$\frac{1}{2}(l_D - l_{D'}) \rightarrow AB + CG + EF$
$\frac{1}{2}(l_E + l_{E'}) \rightarrow E$	$\frac{1}{2}(l_E - l_{E'}) \rightarrow AC + BG + DF$
$\frac{1}{2}(l_F + l_{F'}) \rightarrow F$	$\frac{1}{2}(l_F - l_{F'}) \rightarrow AG + BC + DE$
$\frac{1}{2}(l_G + l_{G'}) \rightarrow G$	$\frac{1}{2}(l_G - l_{G'}) \rightarrow AF + BE + CD$

y los factores a estudiar, y se elige del menú de diseños la fracción deseada. Ahí mismo, el software reporta la resolución que tendrá la fracción elegida. Enseguida se decide sobre el número de repeticiones al centro y se aleatoriza el experimento. Hasta aquí es lo mismo que crear un diseño factorial 2^k completo. Sólo falta decidir de manera específica cuál de las diferentes fracciones posibles se quiere construir.

En el centro de la misma ventana donde se deciden las repeticiones aparece la opción de *Generators*, en la cual se muestran los generadores iniciales de la fracción. Se selecciona una combinación de los signos de los generadores para construir una de las 2^p posibles fracciones, donde p es el número de generadores. La combinación que el *software* sugiere por omisión es la que sólo tiene signos positivos, conocida como *fracción principal*.

Un aspecto importante para decidir la fracción a utilizar en un problema dado es el análisis de la estructura de alias (*Alias structure*) que genera automáticamente el software al construir una fracción. Este análisis permite al experimentador prever el tipo de estructura de confusión que tendrá que interpretar cuando analice los resultados del experimento. Entre las fracciones a elegir también aparecerá el diseño Plackett-Burman.

Una vez realizadas las corridas experimentales y luego de registrar los resultados, se procede con el análisis y con la misma secuencia hasta llegar a Design Analysis.

En Minitab se procede de manera similar, pero siguiendo las opciones: *Stat* → *Doe* → *Factorial* → *Create Factorial Design*. Después se tiene dos opciones: 2-*level factorial* (*default generators*) y 2-*level factorial* (*specify generators*). Con la primera opción se generan las fracciones con los generadores estándar, en la segunda será necesario especificar los generadores deseados. En la pestaña *Display Available Designs* aparece una lista de posibles diseños junto con su resolución.

Preguntas y ejercicios

1. ¿Qué es un diseño factorial fraccionado y cuándo se recomienda aplicarlo?
2. ¿Por qué se dice que un diseño factorial completo con muchos factores (cinco o más) genera un exceso de información?
3. Con el uso del concepto de resolución de un diseño, explique qué se pierde al correr diseños factoriales fraccionados.

4. Considere un diseño factorial fraccionado 2^{5-1}.
   a) ¿Cuántos factores se estudian?
   b) En este diseño sólo se corren 16 de 32 posibles tratamientos, ¿qué se pierde y qué se gana al correr sólo la mitad?
   c) A partir de los 32 tratamientos posibles, ¿se puede seleccionar cualquiera hasta completar 16 o cuál es el criterio de selección?
   d) Elija a $I = ABCDE$ como generador y obtenga la matriz de diseño (los tratamientos que constituyen este diseño).
   e) ¿En qué orden deben correrse los 16 tratamientos del diseño?

5. Con respecto a un diseño factorial 2^4:
   a) Obtenga las dos posibles fracciones a la mitad para este diseño.
   b) ¿Qué resolución tienen estas fracciones y por qué?
   c) ¿Con qué criterio se selecciona una de las dos fracciones?
   d) Obtenga la estructura de alias para ambas fracciones.

6. Considere un diseño factorial fraccionado 2^{8-4}.
   a) ¿Cuántos factores se estudian y cuántas corridas del proceso implica?
   b) ¿Cuántos generadores independientes tiene? ¿Cuántas palabras o generadores tiene la relación definidora?
   c) ¿Cuál es la resolución del diseño? ¿Qué implica esta resolución?

7. ¿Cuándo se dice que una fracción factorial es saturada?

8. Construya un diseño factorial fraccionado 2_{III}^{7-3}.
   a) En términos prácticos, ¿qué significa que el diseño sea de resolución III?
   b) Observe su estructura de alias completa. ¿Cuántos alias tiene cada efecto?
   c) Describa una situación en la que es pertinente utilizar este diseño.

9. Señale las principales características de las fracciones tipo Plackett-Burman y cuándo se recomienda aplicarlas.

10. A continuación se muestran los tratamientos y los resultados obtenidos en un diseño factorial 2^{5-1}.

$(1) = 700$	$de = 2\,515$
$ae = 1\,317$	$ad = 2\,507$
$be = 468$	$bd = 2\,247$
$ab = 424$	$abde = 2\,232$
$ce = 580$	$cd = 2\,031$
$ac = 2\,247$	$acde = 2\,314$
$bc = 446$	$bcde = 2\,262$
$abce = 468$	$abcd = 2\,299$

   a) ¿Cuál es el generador de esta fracción factorial?
   b) ¿Cuál es la resolución de este diseño? ¿Qué significa?
   c) Obtenga un diagrama de Pareto y la gráfica de Daniel para los efectos. ¿Cuáles efectos parecen ser activos?
   d) Obtenga el mejor ANOVA.
   e) Genere la gráfica de los efectos activos en el mejor ANOVA, e interprételos con detalle.
   f) Si lo que se quiere es maximizar, ¿cuál es el mejor tratamiento?

11. Con respecto al problema anterior:
    a) ¿Puede colapsar el diseño factorial fraccionado en un factorial completo? Argumente su respuesta.
    b) Elimine los factores que menos impacto tuvieron, colapse el diseño 2^{5-1} en un diseño 2^3. ¿Cuántas réplicas tiene el diseño colapsado?
    c) Analice con detalle el diseño colapsado y obtenga conclusiones.
    d) ¿Las conclusiones obtenidas con el análisis del diseño fraccionado y el colapsado son las mismas?

12. Una organización de manufactura produce partes de plástico con moldeo por inyección. Típicamente, las partes se encogen de manera excesiva, lo cual causa problemas en el ensamble posterior. Por medio de diseño de experimentos buscan reducir el promedio de encogimiento de las partes y se espera reducir también la variabilidad de corrida a corrida. Se identifican siete factores para el estudio. Cuatro de esos factores son fácilmente controlables: *temperatura del molde* (*A*), *velocidad del tornillo* (*B*), *tiempo de permanencia* (*C*) y *tamaño de compuerta* (*D*). Tres de las variables no son fáciles de controlar durante la manufactura normal: *tiempo de ciclo* (*E*), *contenido de mezcla* (*F*) y *presión en el molde* (*G*), pero durante el experimento se controlarán. Debido a potenciales efectos no lineales en la relación entre el encogimiento y las variables de proceso, se deben considerar al menos tres niveles en cada factor. La alternativa que se considera es utilizar una fracción 2^{7-3} con generadores $I = ABCE$, $I = BCDF$ e $I = ACDG$, que es una octava parte del factorial completo 2^7. Se incluyen cuatro repeticiones al centro para detectar la posible presencia de curvatura. Los datos en unidades de encogimiento × 10 se muestran a continuación.

Orden estándar	Orden aleatorio	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*Y*
1	8	–	–	–	–	–	–	–	6
2	16	+	–	–	–	+	–	+	10
3	18	–	+	–	–	+	+	–	32
4	17	+	+	–	–	–	+	+	60
5	3	–	–	+	–	+	+	+	4
6	5	+	–	+	–	–	+	–	15
7	10	–	+	+	–	–	–	+	26
8	2	+	+	+	–	+	–	–	60
9	9	–	–	–	+	–	+	+	8
10	15	+	–	–	+	+	+	–	12
11	12	–	+	–	+	+	–	+	34
12	6	+	+	–	+	–	–	–	60
13	13	–	–	+	+	+	–	–	16
14	19	+	–	+	+	–	–	+	5
15	11	–	+	+	+	–	+	–	37
16	1	+	+	+	+	+	+	+	52
17	20	0	0	0	0	0	0	0	25
18	4	0	0	0	0	0	0	0	29
19	14	0	0	0	0	0	0	0	24
20	7	0	0	0	0	0	0	0	27

    a) ¿Cuál es la estructura de alias del diseño?
    b) Encuentre el mejor ANOVA para estos datos. No olvide verificar la presencia de curvatura.

c) Proyecte el diseño en uno más simple si hay factores que no afectan, y haga el análisis.

d) Interprete con detalle los efectos activos y determine el mejor tratamiento para reducir el encogimiento promedio.

e) Determine el mejor tratamiento considerando que también interesa reducir la variabilidad.

13. En una empresa panificadora existen problemas con la simetría y el color del pan integral. Los responsables del proceso sospechan que el problema se origina desde la fase de fermentación, en la cual se combina agua, harina, cierta cantidad de levadura más una serie de ingredientes como fosfato, sal, etc. Al final de la fermentación se obtiene lo que llaman "esponja líquida", la cual debe cumplir una serie de parámetros de calidad: una acidez total titulable (ATT) mayor a 6.0 y un pH mayor a 4.8. Deciden utilizar un diseño factorial fraccionado 2^{6-2} para investigar el efecto de seis factores en las variables ATT y pH. Los primeros cinco factores se refieren a cierta cantidad que se agrega en la fermentación: *A*: levadura (17, 19), *B*: sal (2.5, 3.7), *C*: fosfato (2.0, 3.6), *D*: sulfato (1.5, 2.2) y *E*: cloruro (0.89, 1.20); el sexto factor es *F*: temperatura inicial del agua (22, 26). Los datos obtenidos se muestran en la siguiente tabla:

Orden de corrida	Matriz de diseño						Variables de respuesta	
	A	*B*	*C*	*D*	*E*	*F*	ATT	pH
9	−	−	−	−	−	−	6.2	4.86
5	+	−	−	−	+	−	5.6	4.86
6	−	+	−	−	+	+	5.8	4.85
1	+	+	−	−	−	+	5.8	4.99
14	−	−	+	−	+	+	5.7	4.94
10	+	−	+	−	−	+	6.4	4.74
13	−	+	+	−	−	−	6.4	4.83
12	+	+	+	−	+	−	6.6	4.85
11	−	−	−	+	−	+	5.3	4.81
3	+	−	−	+	+	+	6.6	4.81
15	−	+	−	+	+	−	5.2	4.98
16	+	+	−	+	−	−	5.5	4.98
8	−	−	+	+	+	−	6.9	4.84
4	+	−	+	+	−	−	7.1	4.85
2	−	+	+	+	−	+	6.7	4.96
7	+	+	+	+	+	+	6.9	4.84

a) Observe los datos con cuidado, sobre todo los correspondientes al pH. ¿Qué observa de destacado? ¿A qué puede deberse eso?

b) ¿Cuál es la resolución de este diseño y qué significa ésta? Escriba la estructura alias reducida.

c) ¿Cuáles efectos explican el comportamiento de cada una de las respuestas? Encuentre el mejor ANOVA para cada respuesta e interprete utilizando $\alpha = 0.05$.

d) Determine las condiciones de operación que maximizan simultáneamente a ambas respuestas. ¿Es posible dar una solución simultánea al problema con los análisis individuales? Argumente su respuesta.

e) Verifique los supuestos para cada variable de respuesta.

14. Considere un experimento 2^{5-1} con $I = ABCDE$ que fue utilizado para investigar los efectos de cinco factores sobre el color de un producto químico. Los factores son A = solvente/reactante, B = catalizador/reactante, C = temperatura, D = pureza del reactante y E = acidez del reactante. Los resultados obtenidos son los siguientes:

$e = -.63$	$d = 6.79$
$a = 2.51$	$ade = 5.47$
$b = -2.68$	$bde = 3.45$
$abe = 1.66$	$abd = 5.68$
$c = 2.06$	$cde = 5.22$
$ace = 1.22$	$acd = 4.38$
$bce = -2.09$	$bcd = 4.30$
$abc = 1.93$	$abcde = 4.05$

a) Calcule los efectos y grafíquelos en Pareto y en papel normal. ¿Cuáles parecen significativos?
b) Obtenga el mejor análisis de varianza. ¿Con cuáles efectos se está construyendo el error?
c) Represente gráficamente cada efecto significativo e interprételo con detalle.
d) Determine el mejor tratamiento y la respuesta predicha por el modelo.
e) Haga el análisis de residuos y comente los resultados.
f) Si hay algún factor que no tiene ningún efecto, colapse el diseño. ¿Qué diseño resultó?

15. Con el propósito de investigar la precipitación de carbonato de calcio en el interior de una fibra de madera, se decide correr un diseño factorial fraccionado 2^{8-3}_{IV} con los siguientes factores, niveles y variables de respuesta.

Factores		Niveles		Variables de respuesta
		Bajo	Alto	
Tiempo de contacto con primera solución	A	20	35 min.	
Consistencia de $CaCl_2$	B	5	10%	Cenizas Y_1 (%)
Tiempo de contacto con segunda solución	C	20	35 min.	Tensión Y_2 (m)
Agitación	D	1 000	2 000 rpm	Blancuras Y_3 (%)
Concentración de $CaCl_2$	E	40	150 g/L	Opacidad Y_4 (%)
Tiempo de lavado	F	10	20 min.	
Temperatura	G	30	40°C	
Concentración de Na_2CO_3	H	40	150 g/L	

Los resultados obtenidos para los 32 tratamientos corridos se muestran a continuación:

Tratamiento	Y_1	Y_2	Y_3	Y_4	Tratamiento	Y_1	Y_2	Y_3	Y_4
h	7.37	648.1	81.97	81.81	*e*	3.61	347.8	82.41	81.88
afgh	4.52	283.0	86.50	81.60	*aefg*	22.22	179.1	88.75	79.60
bfg	2.17	534.2	83.50	80.49	*befgh*	1.30	436.4	85.40	79.10
ab	3.25	404.0	83.92	85.52	*abeh*	7.73	554.3	84.40	81.16
cf	3.44	602.3	83.50	82.57	*cefh*	1.63	304.2	87.45	80.77
acg	11.99	348.8	87.95	80.00	*acegh*	2.11	876.6	85.75	82.40
bcgh	3.12	317.3	86.57	77.74	*bceg*	23.18	664.6	84.10	82.70
abcfh	0.97	666.7	86.47	78.5	*abcef*	0.79	401.4	89.00	81.97
dg	5.24	503.8	87.12	80.00	*degh*	2.54	427.0	85.30	79.68
adf	3.41	315.0	87.82	81.45	*adefh*	1.35	233.4	85.65	82.63
bdfh	4.72	560.1	87.15	78.14	*bdef*	28.24	721.1	83.60	82.33
abdgh	10.02	322.9	82.73	78.15	*abdeg*	26.32	288.3	88.90	81.75
cdfgh	4.23	350.1	86.95	79.90	*cdefg*	2.24	271.3	88.20	79.20
acdh	4.28	377.8	86.25	79.32	*acde*	6.70	502.3	84.00	81.50
bcd	12.15	638.4	82.48	79.22	*bcdeh*	3.15	293.9	86.25	80.98
abcdfg	6.29	347.8	87.45	80.00	*abcdefgh*	0.31	362.6	88.62	79.60

a) Considere interacciones de dos factores, calcule los efectos y grafíquelos en Pareto y en papel normal para la variable Y_1. ¿Cuáles parecen significativos?
b) Apoyándose en la estructura de alias, ¿alguna de las interacciones dobles que al parecer son importantes, están confundidas?
c) En caso de que la pregunta anterior resulte afirmativa, ¿qué alternativas habría para aclarar la ambigüedad resultante de los efectos que están confundidos? Argumente su respuesta.
d) Interprete con detalle los efectos que parecen ser importantes.
e) Intente obtener el mejor análisis de varianza.
f) Determine el mejor tratamiento si lo que se quiere es maximizar.
g) Si claramente se aprecia que uno o más factores no tiene ningún efecto, colapse el diseño y repita el análisis anterior.
h) Repita los incisos anteriores para las otras variables de respuesta.

16. Ozzie Cadenza, doctor en estadística dueño y gerente de *Ozzie's Bar and Grill*, decidió estudiar los factores que influyen en las ganancias en dólares del bar (véase Box y Draper, 1987). En principio, él no sabía cuáles factores eran importantes y cuáles no lo eran, pero preparó la siguiente lista de *seis* factores, que decidió investigar por medio de un diseño factorial fraccionado. *La cantidad de luz en el bar* (*A*), *el ofrecimiento de papas y chip dip gratis* (*B*), *el volumen de la rockola* (*C*), *la presencia de la cliente favorita del bar* (*D*), una mujer joven de nombre Rapunzel Freeny. La señorita Freeny era el alma de la fiesta en el bar, continuamente "cotorreaba" con los clientes, pasaba las papitas, etc., lo cual hizo pensar a Ozzie que ella tenía un efecto real en la cantidad de ganancias del negocio, *la presencia de la banda musical Gypsy Band* que fueron contratados por Ozzie para tocar un tiempo limitado, y *el efecto del barman que estuviera en servicio* (*E*), había originalmente tres cantineros: Tom, Dick y Harry, pero éste fue despedido para que cada factor en el experimento pudiera tener dos niveles. Signos más y menos fueron asignados a esos seis factores como sigue:

Nivel bajo (–)	Nivel alto (+)
A. Las luces débiles	Las luces altas
B. Sin papas gratis	Con papas gratis
C. La rockola toca bajito	La rockola toca ruidosamente
D. La señorita Freeny en su casa	La señorita Freeny está en el bar
E. Los *gypsies* no están	Los *gypsies* están
F. Tom es el barman	Dick es el barman

Ozzie decidió realizar una corrida cada viernes en la noche durante la hora feliz (4:30 a 6:30 p.m.). Él pensó que debía utilizar un factorial fraccionado con tan pocos tratamientos como fuera posible, puesto que no estaba seguro de en qué momento se podía marchar la banda *Gypsy*. Al final, decidió utilizar un miembro de la familia del factorial 2^{6-3}_{III} con generadores independientes $I = ABD = ACE = BCF$; quería encontrar un diseño de resolución III en el cual nunca ocurriera que la *rockola* tocara a todo volumen al mismo tiempo que los *gypsies*, pero encontró que este requerimiento era imposible.

a) ¿Por qué era imposible?

Él insistió, sin embargo, en que ninguno de los tratamientos que iba a correr tuviera los factores *A*, *C* y *E* simultáneamente en sus niveles altos. Esta restricción era necesaria por la molesta tendencia de las luces a fundirse si los *gypsies* se enfrascaban en su cítara eléctrica al mismo tiempo que las luces estaban altas y la rockola tocaba a todo volumen. Note que esta restricción hizo imposible que se utilizara el miembro principal de la familia de diseños.

b) ¿Cuáles miembros de la familia dada de diseños permiten cumplir esta restricción?

Ozzie estableció la relación generadora $I = ABD = -ACE = BCF$. La matriz de diseño y la ganancia en dólares se muestran en la siguiente tabla.

A	*B*	*C*	*D* = *AB*	*E* = –*AC*	*F* = *BC*	*Y*
–	–	–	+	–	+	265
+	–	–	–	+	+	155
–	+	–	–	–	–	135
+	+	–	+	+	–	205
–	–	+	+	+	–	195
+	–	+	–	–	–	205
–	+	+	–	+	+	125
+	+	+	+	–	+	315

c) Suponiendo que las interacciones de tres o más factores no son importantes, escriba los efectos a estimar.
d) Haga un análisis detallado de los datos obtenidos.
e) ¿Cuáles son las conclusiones más importantes?
f) Dada la estructura de alias, ¿qué dudas razonables podrían surgir sobre las conclusiones?

17. Con respecto al problema de ejercicio anterior, siendo parcial con la señorita Freeny, y alentado por los resultados de la primera fracción, Ozzie escogió una segunda fracción,

la cual daría estimadores limpios del efecto de la señorita Freeny y de todas sus interacciones.

a) Escriba la segunda fracción que logra lo que desea Ozzie. Por cierto, los resultados de esta segunda fracción, dados con las variables *A*, *B* y *C* en orden estándar son: 135, 165, 285, 175, 205, 195, 295, 145.

b) Haga un análisis de los resultados obtenidos con la segunda fracción y obtenga conclusiones.

c) Combine los resultados de ambas fracciones y escriba los efectos estimados.

d) A la luz de esos resultados, ¿fue la selección de la segunda fracción una decisión sabia?

e) Ofrezca una breve conjetura, la cual pueda explicar la presencia y dirección de las interacciones que involucran a la señorita Freeny.

18. En la sección "El concepto de resolución" de este capítulo se mostró que cualquiera de las dos fracciones a la mitad (2^{5-1}) da los mismos resultados que el diseño completo original (2^5) sobre el rendimiento de semiconductores. Muestre, utilizando los datos de la tabla 8.4, que incluso analizando la cuarta parte del experimento (una fracción 2^{5-2}_{III}) es posible llegar a las mismas conclusiones. ¿En qué situación no sería posible llegar a las mismas conclusiones con esta fracción más pequeña?

Capítulo 9
Introducción al diseño robusto (Taguchi)

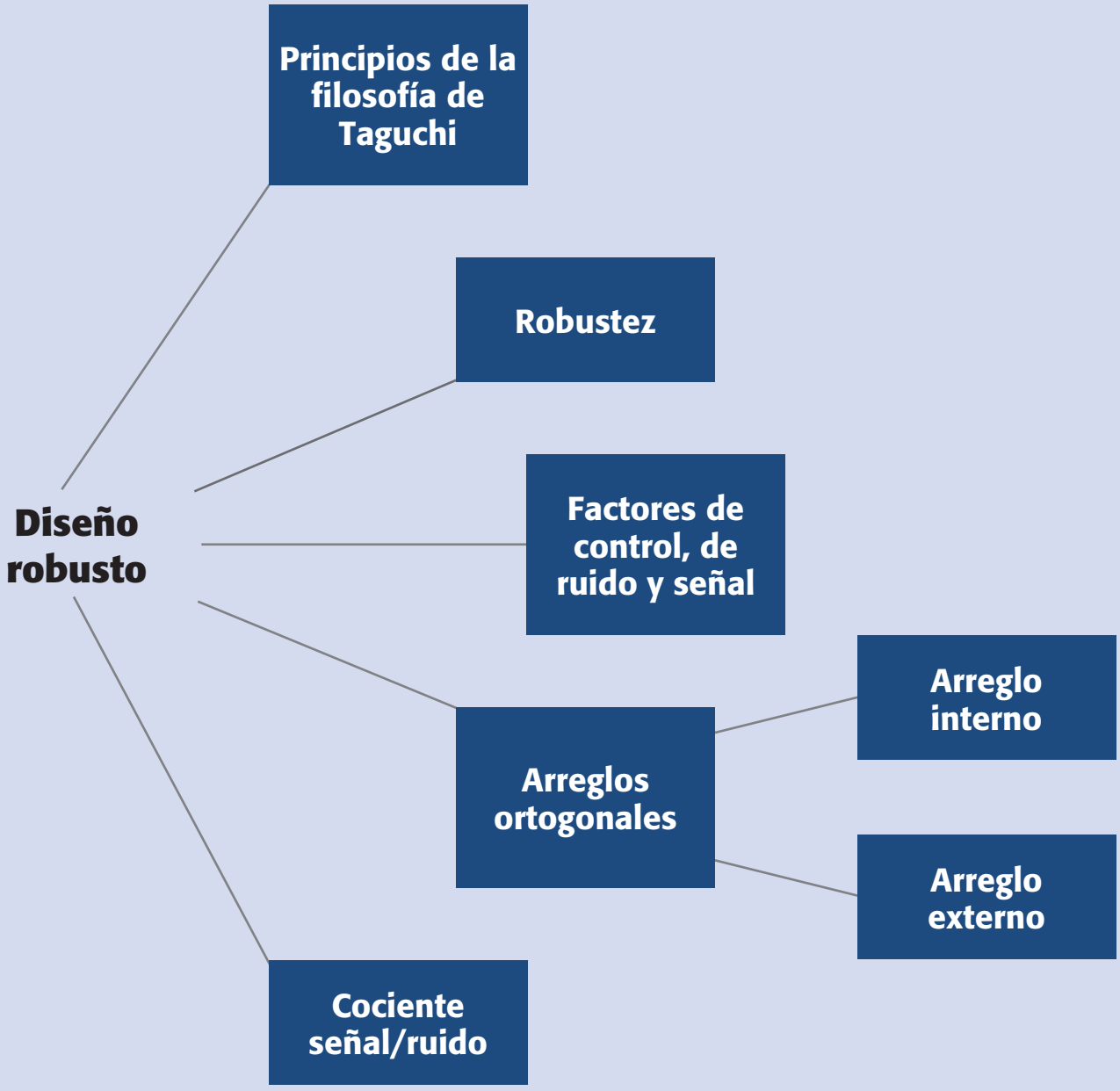
Sumario

- Filosofía Taguchi
- El concepto de robustez
- Factores de control, de ruido y de señal
- Arreglos ortogonales
- Diseño con arreglo interno y externo (diseño de parámetros)
- Razón señal/ruido
- Uso de software

Objetivos de aprendizaje

- Entender los principios de la filosofía Taguchi y el concepto de robustez.
- Identificar los diferentes arreglos ortogonales y a qué tipo de situaciones se aplican.
- Ser capaz de realizar un diseño con arreglo interno y externo, y analizarlo correctamente.

Mapa conceptual



Conceptos clave

- Arreglos ortogonales
- Calidad según Taguchi
- Concepto de robustez
- Deterioro
- Diseño con arreglo interno y externo
- Diseño de parámetros
- Diseño robusto
- Factor de ajuste
- Factor de señal
- Función de pérdida
- Razón señal/ruido
- Ruido externo
- Ruido interno

Diseño de parámetro o diseño robusto
Es la determinación de los niveles de los parámetros o factores de proceso, de tal forma que cada característica del producto se desempeñe con variación mínima alrededor de su valor objetivo.

Calidad según Taguchi
Es la pérdida que un producto causa a la sociedad mientras se utiliza para los fines que fue hecho.

Parámetro
Son los factores o variables del proceso.

El *diseño robusto* tiene su origen en las ideas del ingeniero japonés Genichi Taguchi, quien desarrolló su propia filosofía y métodos de ingeniería de la calidad desde la década de 1950. Fue a partir del éxito de los japoneses en industrias tan importantes como la automotriz y la electrónica que Occidente comienza a fijarse en los métodos utilizados por ellos. De esta manera, los métodos de Taguchi y de otros autores japoneses se introducen en Estados Unidos a partir de la década de los años ochenta.

Taguchi hace varias contribuciones a la calidad. Es precisamente en el diseño de experimentos donde hace sus aportaciones más importantes, con la introducción de lo que él llama *diseño de parámetros* (Taguchi, 1987), que se convirtió en lo que ahora se conoce en Occidente como *diseño robusto*.

En este capítulo describimos las principales ideas del diseño robusto y cómo se pueden utilizar para la mejora y optimización de procesos.

Cabe señalar que la mayoría de las ideas y los métodos propuestos por Taguchi fueron muy criticados por los expertos estadísticos de Occidente (por ejemplo, Nair, 1992 y Box, 1988); no obstante, terminaron por reconocer que Taguchi ha hecho una buena cantidad de contribuciones a las buenas prácticas experimentales que requiere la industria mundial. Quizá la contribución más importante de Taguchi es que, a partir de la controversia generada por sus ideas y métodos, el diseño de experimentos ha logrado un gran avance en las últimas tres décadas, tanto en su base teórica como en aplicaciones a problemas concretos.

Filosofía Taguchi

Taguchi establece que la *calidad de un producto* debe ser medida en términos de abatir al mínimo las pérdidas que ese producto le trae a la sociedad, desde que inicia su fabricación hasta que concluye su ciclo de vida; estas pérdidas sociales se traducen en pérdidas de la empresa en el mediano y largo plazos. Asimismo, se plantea el enfoque al cliente (sociedad) en vez del enfoque al fabricante. Taguchi retoma el concepto del control de calidad fuera de línea (*off line QC*), planteando que la inspección y el control del proceso no son suficientes para alcanzar una calidad competitiva, y que los niveles elevados de calidad sólo pueden lograrse, en términos económicos, en las fases de diseño (producto y proceso).

El objetivo del diseño robusto de parámetros es lograr productos y procesos robustos frente a las causas de la variabilidad (ruidos), que hacen que las características funcionales de los productos se desvíen de sus valores óptimos provocando costos de calidad.

El concepto *parámetro* se refiere a los parámetros del sistema, es decir, son los factores o variables del proceso. Decimos que un producto o proceso es robusto cuando su funcionamiento es consistente al exponerse a las condiciones cambiantes del medio.

La metodología Taguchi establece tres *metas*:

1. Diseños robustos (insensibles) ante el medio ambiente para productos y procesos.
2. Diseño y desarrollo de productos, de modo que sean robustos a la variación de componentes.
3. Minimización de las variaciones con respecto a un valor objetivo.

Estas tres metas se concretan en tres etapas del desarrollo de un producto:

1. *Diseño del sistema*: el ingeniero utiliza principios científicos y de ingeniería para determinar la configuración básica.
2. *Diseño de parámetros*: se determinan los valores específicos para los parámetros del sistema, minimizando la variabilidad aportada por las variables de ruido.
3. *Diseño de tolerancias*: se determinan las mejores tolerancias para los parámetros.

Función de pérdida
Cuantifica la pérdida social que un producto causa cuando sus características de calidad se desvían de su valor ideal.

Un concepto y herramienta clave en el diseño de parámetros es la *función de pérdida*, la cual establece una medida financiera del impacto negativo a la sociedad (consumidor, productor, etc.) por el desempeño de un producto cuando se desvía de un valor designado como meta (t = *target*). Esto implica que la característica de calidad de un producto, y, debe estar cada vez más cerca de su valor ideal, t, y todo lo que se desvíe del ideal es considerado como una pérdida para la sociedad. La función de pérdida de Taguchi (véase figura 9.1), se define como:

$$L(y) = k(y - t)^2$$

donde k es una constante que depende de tolerancias y de los costos de reparación del producto.

De esta ecuación se puede observar que a medida que la característica de calidad y se aleja del valor ideal t, la pérdida aumenta. De esta manera, los esfuerzos de mejora deben estar orientados a reducir la variabilidad de y en torno al valor ideal t, con lo que la pérdida será cada vez más pequeña. Esto contrasta con el pensamiento tradicional que sólo penaliza si y está fuera de especificaciones.

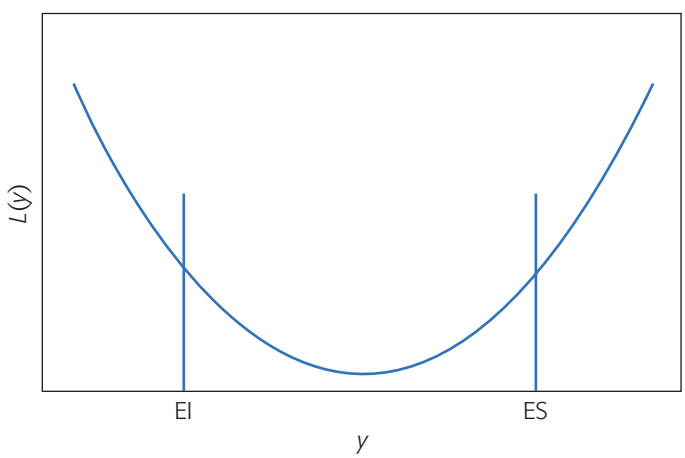


Figura 9.1 Forma típica de la función de pérdida de Taguchi.

El concepto de robustez

Diseño robusto
Experimento en el que se consideran factores de ruido, con los cuales se quiere lograr un proceso o producto robusto.

Robustez
Es hacer un producto o proceso insensible o resistente a factores de ruido que no son controlables.

Un *diseño robusto* es un experimento en el cual existen factores de ruido (no controlables), considerados de manera explícita o implícita, cuyo efecto se pretende minimizar de forma indirecta (o sea sin controlarlo directamente), a fin de encontrar la combinación de niveles de los factores de proceso que sí se pueden controlar, y en donde el efecto de dichos factores de ruido es mínimo. Dicho de otra manera, en un experimento robusto se trata de lograr que el producto/proceso tenga el desempeño deseado sin que le afecten las fuentes de variación no controladas. El objetivo fundamental de un diseño robusto es determinar la combinación de niveles de los factores controlables, en donde los factores de ruido no afecten al proceso, aunque estos últimos no se controlen. El significado de la palabra *robusto* es en el sentido de hacer el proceso o producto insensible o resistente a factores de ruido que no está en nuestras manos controlar. A continuación ilustraremos el concepto de robustez con algunos ejemplos sobre el diseño de artículos de uso común.

Ejemplo 9.1

Consideremos una copiadora. Al momento de sacar copias, el usuario desea que la máquina funcione bien sin importar el tipo de papel usado ni la humedad ambiental, por mencionar sólo dos factores de ruido que el fabricante de las copiadoras no puede controlar. Si la máquina trabaja bien en esas condiciones es posible afirmar que es robusta al papel y a la humedad ambiental. Además, el usuario también desea elegir entre una reducción o un aumento en el tamaño de la copia, así como lo oscuro de la tinta, y que la copia siempre salga con la calidad deseada.

Ejemplo 9.2

Si pensamos en una tostadora de pan, al usuario le gustaría que fuera robusta a la marca de pan, a la humedad ambiental, a la temperatura ambiental, a las variaciones en las dimensiones del pan, etc. Se desea seleccionar el grado de tostado y observar que, efectivamente, el pan salga con el color deseado, en toda su superficie y por ambos lados.

Ejemplo 9.3

En una fábrica de dulces se tenía el problema de que la plasticidad del caramelo era altamente dependiente de la temperatura ambiental, de manera que cuando hacía mucho calor se escurría sobre las manos del consumidor final. Entonces, el problema era formular un dulce robusto a la temperatura ambiental. Se encontró, mediante un experimento robusto, una nueva formulación de caramelo más resistente a los cambios de temperatura. El efecto de la temperatura ambiental sobre la plasticidad del caramelo, antes y después de la mejora, se muestra en la figura 9.2.

El diseño robusto se enfoca a la fabricación de productos y procesos robustos, lo cual se logra mejor durante la etapa en que se concibe y diseña un nuevo producto; además, en esta etapa es posible reducir el costo al incluir materiales más económicos que cumplan la función deseada. Tener un proceso robusto significa que éste funcione bien aunque varíen una serie de factores (de ruido) que no se pueden controlar, como variables ambientales (temperatura, humedad, etc.), cansancio de los

operadores, cambios de turno y de lotes, variaciones no controlables en variables de proceso, acumulación de suciedad, etcétera.

Factores de control, de ruido y de señal

Como se mencionó en los capítulos previos, en un proceso existen básicamente dos tipos de factores: controlables y no controlables (o de ruido). Por lo regular, los factores de estudio eran controlables. Sin embargo, en diseño robusto es conveniente tener una clasificación más detallada del tipo de factores controlables que pueden influenciar el proceso, en cuanto a su efecto sobre la media y la variabilidad de la respuesta de interés. Se distinguen cuatro tipos de factores, a saber:

I. Afecta la media y la variabilidad.
II. Afecta sólo la variabilidad.
III. Afecta sólo la media.
IV. No afecta la media ni la variabilidad.

Cuando en el diseño clásico se afirma que un factor tiene efecto sobre la respuesta, por lo general se quiere decir que el factor tiene efecto sobre la media de la característica de calidad. En cambio, en el diseño robusto se tiene más presente que el efecto de un factor también puede ser sobre la variabilidad, o sobre la media y la variabilidad de manera simultánea. Estos cuatro casos se representan en la figura 9.3, en la que en el eje horizontal se muestra el efecto sobre la media y en el eje vertical el efecto sobre la variabilidad. El factor clase I afecta a la media y a la variabilidad, el clase II sólo a la variabilidad, el clase III afecta sólo a la media y el clase IV no afecta a ninguna de las dos (figura 9.3*a*, *b*, *c* y *d*).

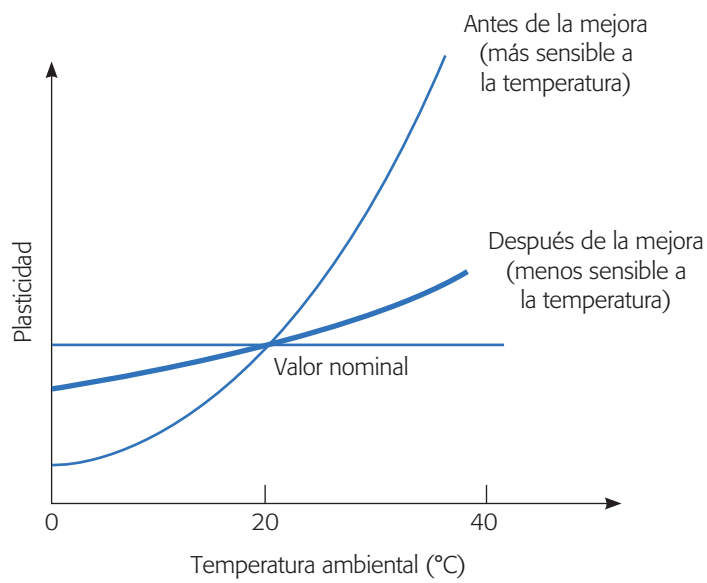


Figura 9.2 Efecto de la temperatura antes y después de la mejora de un caramelo, ejemplo 9.3.

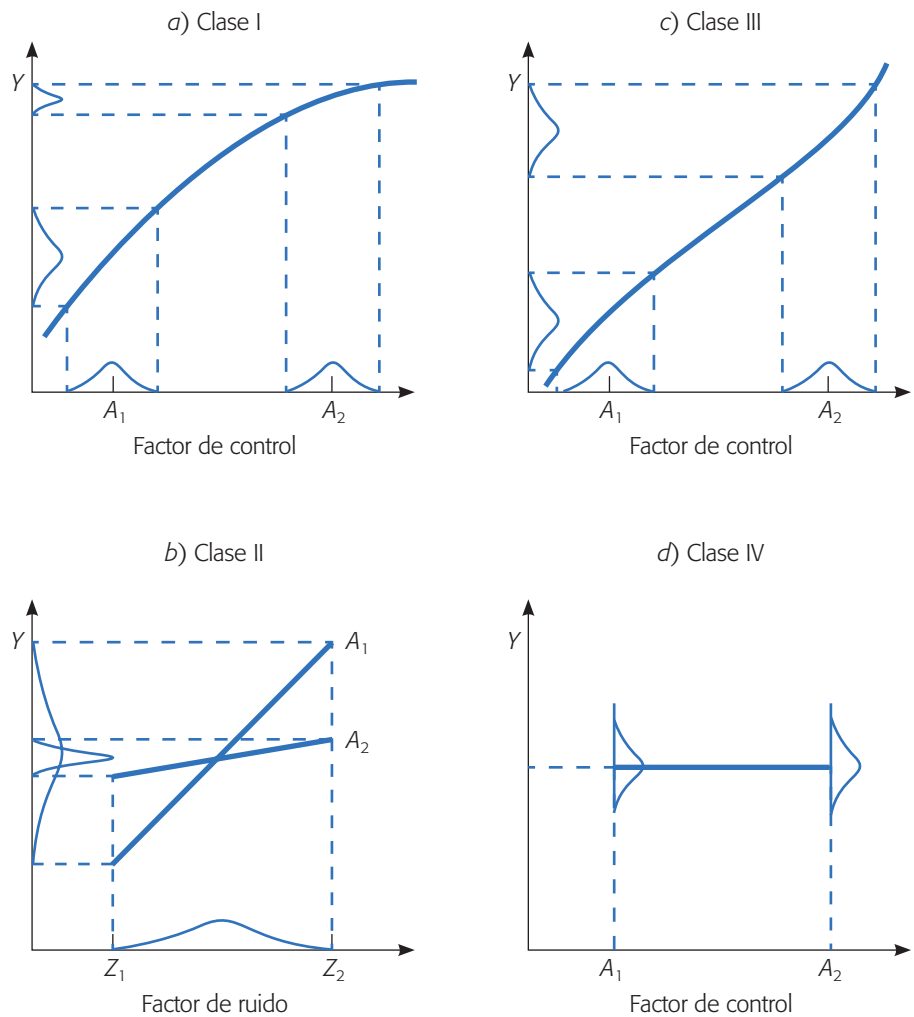


Figura 9.3 Clases de factores de control de acuerdo a su efecto sobre la media (eje *X*) y/o la variabilidad (eje *Y*).

El factor clase I tiene una relación curva con la variable de respuesta. Considerando dos niveles (A_1 y A_2) de este factor, es claro de la gráfica que una misma oscilación o variación del factor sobre cada uno de estos niveles (representada por las curvas sobre ellos) tiene un efecto distinto sobre la variable de respuesta. En el nivel A_2 la respuesta Y tendría una variabilidad menor que en el nivel A_1, es decir, en el nivel A_2 se tiene un comportamiento más robusto del proceso a las posibles oscilaciones del factor. Es claro que al mismo tiempo en A_2, la media de la característica es mayor que en A_1.

El factor de control clase II se representa interactuando con un factor de ruido. Esta interacción tiene efecto principalmente en la variabilidad, ya que el factor de ruido no se controla y seguirá variando entre sus valores extremos Z_1 y Z_2, de manera que en cualquiera de los niveles (A_1 y A_2) la media de la característica es la misma. Lo más relevante de esta interacción es que en el nivel A_1 del factor de control se

minimiza el efecto del factor de ruido sobre la variabilidad de Y. Entonces, se puede afirmar que en el nivel A_1 el proceso es más robusto al efecto del factor de ruido. Este tipo de relación entre un factor de control y un factor de ruido es la idea fundamental que permite hacer diseño robusto. Si en un experimento dado los factores de control actuaran independientemente de los factores de ruido sería imposible hacer diseño robusto, ya que no existiría una combinación de los primeros que minimizara el efecto de los segundos.

El factor clase III afecta sólo la media de la característica de calidad. La variabilidad que pasaría a la respuesta (Y) por la oscilación de este factor es la misma en cualquier lugar de su rango. Este tipo de factor es útil para llevar la media a su valor nominal una vez elegidos los niveles de los factores clases I y II que minimizan la variabilidad, de aquí que al factor clase III se le llame *factor de ajuste.* Esto es, con los factores clases I y II es posible elegir las condiciones más robustas, reduciendo la variabilidad de la respuesta, pero la media se habrá movido de su valor deseable; con el factor clase III ésta se regresa a su valor nominal sin afectar la variabilidad. Finalmente, el factor clase IV no tiene efecto ni en la media ni en la variabilidad y de éste se elige su nivel más económico como el mejor.

Factor de ajuste
Es el factor clase III que afecta sólo la media de la característica de calidad.

En la medida que se conoce la relación entre los factores controlables y no controlables con la variable de respuesta, se está en posición de establecer mejores condiciones de operación del proceso. En el diseño robusto se trata de sacar ventaja principalmente de los factores de control clase II que interactúan con factores de ruido: se trata de elegir el nivel del factor controlable que hace al proceso más insensible al ruido. Después, se busca ajustar la media al valor nominal con un factor de ajuste (clase III).

Factor señal

Muchos productos están diseñados para trabajar en diferentes niveles de desempeño y de acuerdo a los deseos del usuario o consumidor. En otras palabras, el usuario puede elegir la señal (o valor promedio de la respuesta) que desea en un momento dado del producto. Se llama *factor señal* al dispositivo que permite cambiar el nivel de operación de acuerdo a los deseos del usuario. Por ejemplo, en una tostadora de pan el factor señal es el mecanismo que permite seleccionar el grado de tostado deseado; en el caso de una copiadora, un factor señal es el mecanismo para elegir la oscuridad deseada de la impresión. En el caso del limpiaparabrisas de automóvil, el factor señal son las diferentes velocidades que el conductor puede elegir de acuerdo al clima imperante.

Factor señal
Factor de ajuste cuyo nivel de operación es seleccionado por el usuario.

En cualquier caso, se trata de lograr que el dispositivo proporcione la señal o tenga el desempeño con la calidad que el usuario espera. Note que el factor señal permite cambiar el valor de la media de la característica de calidad, y es deseable que la variabilidad en cada nivel de operación sea mínima. Es decir, el producto debe ser robusto en cada nivel del factor señal. Por ejemplo, el factor de control clase III de la figura 9.3, que sólo afecta la media, podría hacer las veces de factor señal, puesto que permitiría modificar el valor de la media sin que cambie la variabilidad de la respuesta.

Cabe aclarar que hay muchos productos que no cuentan con un dispositivo para que el usuario elija cierto nivel de desempeño, y en ese caso no existe tal factor señal.

Por ejemplo, si una aspiradora sólo cuenta con una velocidad, no tiene factor señal por este concepto. En el otro extremo, existen productos que tienen más de un factor señal que el usuario puede manipular de manera independiente. Por ejemplo, en el caso de la copiadora se tienen dos factores señal, puesto que el usuario puede seleccionar la oscuridad de la impresión y el tamaño de la misma, y en cualquier combinación de ellas espera un resultado de calidad.

Factores de ruido

Ruido externo
Ambiente en el cual el proceso o producto se desempeña, y carga de trabajo a que se somete.

Ruido interno
Variación generada por el proceso de unidad a unidad debido a su tecnología y a la diversidad de sus componentes.

Deterioro
Efectos que aparecen poco a poco con el tiempo por la degradación del proceso, y pueden llegar a causar fallas.

Los factores de ruido que actúan sobre el producto o sobre el proceso se clasifican como: de ruido externo, ruido interno y de deterioro. El *ruido externo* se refiere al ambiente en el cual el proceso (o producto) se desempeña y a la carga de trabajo a que es sometido. Por ejemplo, es ruido externo la humedad ambiental, el polvo o los errores en la operación del equipo. El *ruido interno* se refiere a la variación generada por el proceso de unidad a unidad producida, y que se debe a su propia naturaleza o tecnologías y la diversidad de sus componentes. El *deterioro* se refiere a efectos que aparecen poco a poco con el tiempo por la degradación paulatina del proceso y sus componentes, que pueden causar la aparición de fallas en el proceso/producto. Por ejemplo, piezas o herramientas que se van gastando por el mismo uso.

En la figura 9.5 se muestran los diferentes tipos de factores que intervienen en el diseño robusto. Lo nuevo en este diagrama en relación a los del capítulo 1 es el factor señal que se acaba de definir.

Tipos de estudios de robustez

Los estudios de robustez se clasifican utilizando como criterios al tipo de variable de respuesta y la existencia o ausencia de factores de señal. Una variable de respuesta puede ser de tres tipos:[1] entre más pequeña mejor, entre más grande mejor o nominal es lo mejor. En cuanto al factor señal, se dice que el estudio es estático si no hay factor señal y es dinámico en el otro caso.

Ejemplo 9.4
Experimento robusto: ejemplo integrador

Una de las características importantes en el proceso de producción de un pigmento es su color. El problema que se tenía en este proceso era el exceso de variación del color del pigmento. Un grupo de mejora decide utilizar diseño robusto para tratar de

[1] *Entre más pequeña mejor.* Son variables o características de calidad cuya única exigencia es que no excedan cierto valor máximo tolerado o especificación superior (ES), y entre más pequeño sea su valor mejor. Por ejemplo: porcentaje de impurezas en una sustancia o la cantidad de sustancias tóxicas en un producto alimenticio.

Entre más grande mejor. Son variables o características de calidad a las que se les exige que sean mayores que un valor mínimo o que cierta especificación inferior (EI), y entre más grande sea el valor de la variable es mejor. Por ejemplo, la resistencia de una pieza de plástico inyectado o la "blancura" de una tela de color blanco.

Valor nominal es el mejor. Variables que deben tener un valor específico y que, por lo tanto, no deben ser menores que una especificación inferior (EI), pero tampoco mayores que una superior (ES). Ejemplos de este tipo de características de calidad con doble especificación son el diámetro interior de una tuerca y la longitud de una pieza para ensamble.

hacer el proceso menos sensible al efecto de factores de ruido difíciles de controlar durante la producción. Se identificaron seis factores de control y tres de ruido con dos niveles cada uno: (1, 2), los cuales se muestran en la tabla 9.1. Se decide utilizar un arreglo ortogonal L_8 para los factores de control y un L_4 para los factores de ruido, con lo que el diseño resultante tiene 32 corridas (pruebas) a nivel proceso. El diseño y los datos obtenidos se muestran en la figura 9.4.

Se procede a calcular los estadísticos de interés en cada combinación de niveles del arreglo interno o de factores de control. Los valores de la media, desviación estándar y del estadístico señal/ruido nominal tipo II se muestran en las tres últimas columnas de la figura 9.4. A continuación explicamos de manera detallada los conceptos nuevos que están involucrados en este ejemplo, como son: los arreglos ortogonales, los arreglos interno y externo y la razón señal/ruido, para así entender el diseño de la figura 9.4 y poder analizarlo.

Arreglos ortogonales

Arreglos ortogonales
Matrices de diseños factoriales completos, fraccionados o mixtos que tienen la propiedad de ortogonalidad.

Los *arreglos ortogonales* son diseños propuestos por Taguchi que, como su nombre lo indica, tienen la propiedad de ortogonalidad,[2] misma que también poseen los dise-

Tabla 9.1 Factores de control y de ruido en la producción de un pigmento.

Factores de control	Factores de ruido
A: Tiempo de carga de materiales	*K*: Calidad de la sal
B: Tiempo de amasado	*L*: Aspecto de la resina
C: Exceso de sal	*M*: Temperatura del agua para enfriar
D: Temperatura de amasado	
E: Orden de introducción de materiales	
F: Velocidad de amasado	

							Factores de ruido						
							K	*L*	*M*	2 1 1	1 1 2	1 2 1	2 2 2	Media	Desviación estándar	Razón señal/ ruido
Factores controlables																
A	*B*	*C*	*D*	*E*	*F*	*G*								$\bar{X}$	*S*	$-10\log_{10}(S^2)$
1	1	1	1	1	1	1				36	26	24	15	25.25	8.61	−18.71
1	1	1	2	2	2	2				32	62	24	32	37.50	16.76	−24.49
1	2	2	1	1	2	2				34	16	25	12	21.75	9.81	−19.83
1	2	2	2	2	1	1				10	30	26	32	24.50	9.98	−19.98
2	1	2	1	2	1	2				33	31	27	23	28.50	4.43	−12.93
2	1	2	2	1	2	1				34	48	26	39	36.75	9.21	−19.29
2	2	1	1	2	2	1				26	27	18	20	22.75	4.42	−12.92
2	2	1	2	1	1	2				28	40	21	32	30.25	7.93	−17.99

Figura 9.4 Diseño con arreglos interno y externo para hacer más robusto el color del pigmento.

[2] Se dice que una matriz de diseño es ortogonal si sus columnas son linealmente independientes, lo cual se tiene si la multiplicación de dos columnas cualesquiera es igual a cero.

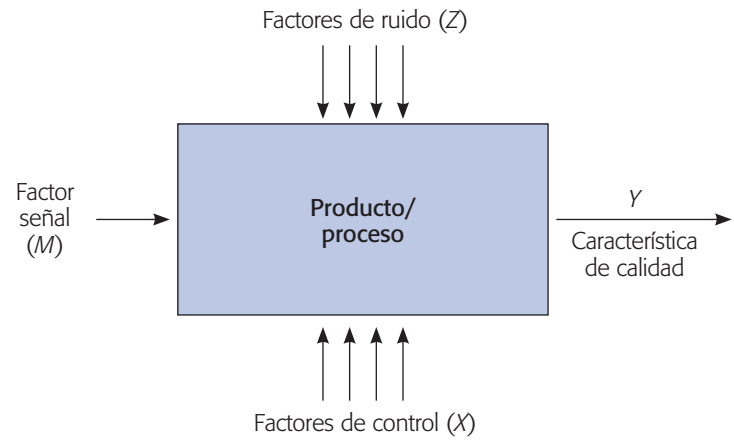


Figura 9.5 Diagrama de los diferentes tipos de factores en diseño robusto.

ños factoriales clásicos. Estos arreglos son diseños factoriales completos, fraccionados o mixtos, dependiendo del número de factores a estudiar en un caso particular. Por ejemplo, el arreglo ortogonal L_8 (AO_L_8) tiene ocho corridas experimentales, y con él se pueden estudiar desde dos hasta siete factores en dos niveles cada uno (figura 9.6). Si se estudian siete factores equivale a un diseño factorial fraccionado 2^{7-3}_{III}, mientras que con dos factores sería un factorial completo 2^2 con dos réplicas. Para menos de siete factores es necesario saber a cuáles columnas del arreglo deben asignarse los factores, para tener la estructura alias que permita estudiar lo más claro posible a esos factores. Por ejemplo, al estudiar cuatro factores con el arreglo L_8, éstos se deben asignar a las columnas 1, 2, 4 y 7, para obtener la información con la calidad de la fracción 2^{4-1}_{IV} con generador $I = ABCD$ (capítulo 8).

En el ejemplo 9.4 del color del pigmento, en la parte interior de la figura 9.4, se aplica un arreglo interno L_8 para decidir qué combinación de los seis factores controlables se correrán. Las columnas se asignaron a los factores en forma consecutiva y no en la forma que se recomienda, por lo que la fracción resultante quizá no tenga la máxima resolución.

Mientras que el arreglo externo que indica las combinaciones de los tres factores de ruido, se seleccionaron de acuerdo a un arreglo L_4 (véase parte superior de la figura 9.4), pero sustituyendo el nivel "1" por el "2" y el "2" por el "1".

Con el arreglo ortogonal L_{16} se pueden estudiar desde cuatro hasta 15 factores (para menos de cuatro factores es mejor utilizar un L_8). Cuando se estudian cuatro factores, éstos se deben asignar a las columnas 1, 2, 4 y 8, para estimar de manera separada a las interacciones, y en este caso equivaldría a un diseño factorial completo 2^4. Si se estudian 15 factores el arreglo L_{16} equivaldría a una fracción 2^{15-11}_{III}. Ocurre que, en la medida de que se estudian más factores, se tienen menos grados de libertad para estudiar interacciones; sin embargo, Taguchi no hace énfasis en el estudio de las interacciones, prefiere saturar lo más posible los arreglos y analizar sólo los efectos principales de cada factor.

En la figura 9.6 se muestran seis de los arreglos ortogonales de uso más frecuente, que son: L_4, L_8, L_{12}, L_{16}, L_9 y L_{18}, y en la parte inferior de cada uno de ellos

Arreglo L_4 (fracción 2^{3-1})

Núm. de corrida	Núm. de columna		
	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

2 factores: columas 1 y 2.
3 factores: las tres columnas.

Arreglo L_9 (3^{4-2})

Núm. de corrida	Núm. de columna			
	1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

2 factores: columas 1, 2.
3 factores: columnas 1, 2, 3.
4 factores: columnas 1, 2, 3, 4.

Arreglo L_8 (fracción 2^{7-4})

Núm. de corrida	Núm. de columna						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

2 factores: columas 1, 2.
3 factores: columnas 1, 2, 4.
4 factores: columnas 1, 2, 4, 7.
5 factores: columnas 1, 2, 4, 7, 6.
6 factores: columnas 1, 2, 4, 7, 6, 5.
7 factores: las siete columnas.

Arreglo L_{12} (Plackett-Burman para $k = 11$)

Núm. de corrida	Núm. de columna										
	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	2	2	2	2	2	2
3	1	1	2	2	2	1	1	1	2	2	2
4	1	2	1	2	2	1	2	2	1	1	2
5	1	2	2	1	1	2	1	2	1	2	1
6	1	2	2	2	2	2	2	1	2	1	1
7	2	1	2	2	2	1	2	2	1	2	1
8	2	1	2	1	1	2	2	1	1	1	2
9	2	1	1	2	2	2	1	2	2	1	1
10	2	2	2	1	1	1	1	2	2	1	2
11	2	2	1	2	2	2	1	1	1	2	2
12	2	2	1	1	1	1	2	1	2	2	1

Se asignan los k factores a las primeras k columnas ($4 < k < 11$).

Arreglo L_{18} ($2 \times 3^{7-5}$)

Núm. de corrida	Núm. de columna							
	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1
2	1	1	2	2	2	2	2	2
3	1	1	3	3	3	3	3	3
4	1	2	1	1	2	2	3	3
5	1	2	2	2	3	3	1	1
6	1	2	3	3	1	1	2	2
7	1	3	1	2	1	3	2	3
8	1	3	2	3	2	1	3	1
9	1	3	3	1	3	2	1	2
10	2	1	1	2	3	2	2	1
11	2	1	2	1	1	3	3	2
12	2	1	3	2	2	1	1	3
13	2	2	1	2	3	1	3	2
14	2	2	2	3	1	2	1	3
15	2	2	3	1	2	3	2	1
16	2	3	1	3	2	3	1	2
17	2	3	2	1	3	1	2	3
18	2	3	3	2	1	2	3	1

1 factor con dos niveles se asignan a la columna 1.
Los factores con tres niveles se asignan a las columnas restantes: 2, 3, 4, 5, 6, 7, 8.

Figura 9.6 Arreglos ortogonales más frecuentes.

Arreglo L_{16} (2^{15-11})															
Núm. de corrida	Núm. de columna														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
5	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1
7	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1

4 factores: columnas 1, 2, 4, 8
5 factores: columnas 1, 2, 4, 8, 15
6 factores: columnas 1, 2, 4, 8, 15, 14
7 factores: columnas 1, 2, 4, 8, 15, 14, 13
8 factores: columnas 1, 2, 4, 8, 15, 14, 13, 11
9 factores: columnas 1, 2, 4, 8, 15, 14, 13, 11, 7
10 factores: columnas 1, 2, 4, 8, 15, 14, 13, 11, 7, 12
11 factores: columnas 1, 2, 4, 8, 15, 14, 13, 11, 7, 12, 10
12 factores: columnas 1, 2, 4, 8, 15, 14, 13, 11, 7, 12, 10, 9
13 factores: columnas 1, 2, 4, 8, 15, 14, 13, 11, 7, 12, 10, 9, 6
14 factores: columnas 1, 2, 4, 8, 15, 14, 13, 11, 7, 12, 10, 9, 6, 5
15 factores: columnas 1, 2, 4, 8, 15, 14, 13, 11, 7, 12, 10, 9, 6, 5, 3

Figura 9.6 Arreglos ortogonales más frecuentes. (*Continuación*.)

se lista la manera de asignar los factores a las columnas, lo cual es importante si no se van a ocupar todas, ya que una buena asignación evita confundir los efectos principales o incluso separar algunos efectos de interacción. El subíndice en la notación L_i indica el número de combinaciones de niveles que conforman el arreglo. Los arreglos L_9 y L_{18} permiten estudiar factores con tres niveles (1, 2, 3).

Note que Taguchi acomoda las columnas de los arreglos ortogonales en un orden diferente al orden de Yates que se estudió en los diseños factoriales. La primera columna de cada arreglo ortogonal es aquella donde los niveles aparecen lo más agrupados posible, de manera que el factor correspondiente se cambia de nivel un número mínimo de veces si el arreglo se corre en este orden. Taguchi recomienda asignar a la primera columna aquel factor que sea más difícil de manipular durante el experimento; es decir, el factor al que sea difícil cambiarle su nivel de una prueba a otra. De lo anterior se concluye que Taguchi no enfatiza la necesidad de correr el experimento en orden aleatorio como se recomienda en diseño clásico, sino más bien presupone las complicaciones prácticas que se han señalado, y estructura el orden de las columnas conforme a esas dificultades.

Diseño con arreglo interno y externo (diseño de parámetros)

La condición fundamental para que un diseño experimental sea de tipo robusto es que exista al menos un factor de ruido para el cual se busca hacer que el proceso o producto sea insensible a su efecto, sin pretender controlar dicho factor de ruido. Éste seguirá actuando como siempre el proceso después del experimento, pero se busca que su efecto sea menor. Un diseño experimental propuesto por Taguchi para determinar condiciones de operación robustas a uno o varios factores de ruido es el *diseño con arreglo interno y externo*. Una vez identificados los factores de control y los factores de ruido con los que se quiere experimentar, se construyen dos arreglos ortogonales, uno para cada tipo de factores como se ilustra en la figura 9.4 para el ejemplo del color del pigmento, donde se utiliza un arreglo interno L_8 para determinar las combinaciones de los seis factores de control (factorial fraccionado 2_{III}^{6-3}) y un arreglo L_4 para los factores de ruido. El diseño resultante tiene 32 corridas, y consiste en sobreponer ambos arreglos de manera que en cada combinación de los factores controlables se prueben todas las combinaciones de los factores de ruido (figura 9.4). Note que en cada combinación de los factores controlables (arreglo interno) se están "simulando" distintos tipos de ruido que pueden ocurrir (arreglo externo). La mejor combinación de los factores de control es aquella donde los ruidos tienen el menor efecto (causan mínima variación) y, al mismo tiempo, la media del color se encuentra más cerca del valor deseado.

Diseño con arreglo interno y externo
Sirve para determinar condiciones de operación robustas a uno o varios factores de ruido. Consiste en probar todas las combinaciones de los factores de ruido en cada combinación de los factores de control.

En la figura 9.7 se muestra un diseño con el arreglo interno L_8 (factorial fraccionado 2_{III}^{7-4}) y como arreglo externo L_9 (factorial fraccionado 3_{III}^{4-2}), cada uno de ellos con la cantidad máxima de factores. Esto es, se tienen siete factores de control y cuatro de ruido. El diseño completo consta de 72 corridas experimentales. Una desventaja del diseño con arreglo interno y externo es que requiere una cantidad grande de corridas experimentales, aun utilizando los arreglos ortogonales más pequeños. De aquí que algunos autores (Grize, 1995) propongan como alternativa (con menos corridas) utilizar diseños factoriales completos o fraccionados clásicos incorporando los factores de ruido como factores de control, y cuidando que el diseño permita estudiar la posible interacción entre unos y otros. Al interpretar dichas interacciones se elige el nivel del factor de control en el cual el impacto negativo del factor de ruido sea menor.

Razón señal/ruido

Para el análisis del diseño con arreglo interno y externo, Taguchi propone un estadístico de desempeño, al cual le llama *cociente* o *razón señal/ruido* (*signal to noise ratio*), que se calcula en cada combinación de los factores controlables (figuras 9.4 y 9.7) y se analiza como cualquier variable de respuesta. La combinación más robusta de los niveles de los factores controlables es aquella que maximiza el estadístico razón señal/ruido.

Razón señal/ruido
Estadístico que mide la robustez en cada combinación de los factores de control. Siempre se busca maximizarlo.

De acuerdo con el tipo de característica de calidad, el estadístico razón señal/ruido se define de acuerdo con la tabla 9.2.

								Arreglo externo										
							N	1	2	3	3	1	2	2	3	1		
						Factores de ruido	M	1	2	3	2	3	1	3	1	2		
							L	1	2	3	1	2	3	1	2	3		
							K	1	1	1	2	2	2	3	3	3		
	Factores controlables																	
	A	B	C	D	E	F	G										Media	Razón señal/ruido
Arreglo interno	1	1	1	1	1	1	1	y_{11}	y_{12}	•	•	•				y_{19}	$\bar{y}_{1\cdot}$	S/R_1
	1	1	1	2	2	2	2	y_{21}	y_{22}	•	•	•				y_{29}	$\bar{y}_{2\cdot}$	S/R_2
	1	2	2	1	1	2	2	•	•							•	$\bar{y}_{3\cdot}$	S/R_3
	1	2	2	2	2	1	1	•	•							•	$\bar{y}_{4\cdot}$	S/R_4
	2	1	2	1	2	1	2	•	•							•	$\bar{y}_{5\cdot}$	S/R_5
	2	1	2	2	1	2	1										$\bar{y}_{6\cdot}$	S/R_6
	2	2	1	1	2	2	1										$\bar{y}_{7\cdot}$	S/R_7
	2	2	1	2	1	1	2	y_{81}	y_{82}	•	•	•				y_{89}	$\bar{y}_{8\cdot}$	S/R_8

Figura 9.7 Ejemplo de diseño con arreglo interno (L_8) y arreglo externo (L_9).

El hecho de que se saque logaritmo en los estadísticos tiene que ver con buscar que los efectos de los factores controlables sean *aditivos*, es decir, que se minimice la posibilidad de efectos de interacción entre ellos. Se multiplican por 10 para trabajar en una escala más grande y el signo que lo antecede se escoge de manera que el problema siempre sea maximizar el valor del estadístico para obtener las condiciones de operación más robustas.

En cuanto a los dos estadísticos para respuestas del tipo nominal, se recomienda el tipo I para respuestas no negativas que van de cero a infinito, que tienen valor

Tabla 9.2 Razones señal/ruido para los diferentes tipos de variables de respuesta.

Tipo de característica	Razón señal/ruido (S/R)
Mientras más pequeña es mejor	$-10 \log\left[\frac{1}{n}\Sigma_{i=1}^{n} Y_i^2\right]$
Mientras más grande es mejor	$-10 \log\left[\frac{1}{n}\Sigma_{i=1}^{n} \frac{1}{Y_i^2}\right]$
Su valor nominal es lo mejor (tipo I)	$10 \log\left(\frac{\bar{Y}^2}{S^2}\right)$
Su valor nominal es lo mejor (tipo II)	$-10 \log (S^2)$
Proporción de defectuosos	$-10 \log\left(\frac{p}{(1-p)}\right)$

objetivo diferente de cero y que la varianza es cero cuando la respuesta también es cero. Por su parte, el estadístico tipo II es para respuestas que pueden tomar tanto valores positivos como negativos, y donde el cero puede ser el valor nominal. El estadístico tipo I también se recomienda cuando la media y la desviación estándar tienen una relación lineal; y el estadístico tipo II cuando la media y la desviación estándar son independientes (Fowlkes y Creveling, 1995). Para verificar el tipo de relación entre la media y la desviación estándar se realiza un diagrama de dispersión con los puntos $(\bar{Y}_i, S_i)$ calculados en los renglones del arreglo interno. Si los puntos caen en una banda horizontal a lo largo del eje X, entonces la media y la desviación estándar son independientes. Pero si los puntos muestran alguna relación lineal, o de otro tipo, entonces son dependientes.

Los estadísticos señal/ruido son algunos de los aspectos del método propuesto por Taguchi que más polémica generaron (Box, 1988). Se llegó a demostrar que tales estadísticos ofrecen resultados subóptimos, ya que confunden el efecto sobre la media con el efecto en la variabilidad. Una alternativa que se propuso es transformar los datos de manera que se logre un comportamiento independiente de la media y la variabilidad, y luego analizarlas de manera separada para los datos transformados (Nair y Pregibon, 1986). Otra alternativa más directa y práctica, que en muchos casos da buenos resultados, es utilizar con cualquier tipo de respuesta continua el estadístico del caso nominal, lo mejor es emplear el tipo II ($-10 \log(S^2)$).

Optimización en dos pasos

La estrategia de análisis del experimento con arreglos interno y externo se resume en los dos pasos siguientes:

1. Se determinan los factores controlables que tienen efecto sobre el estadístico razón señal/ruido (S/R), que fue seleccionado de acuerdo con el tipo de característica de calidad que se tiene (véase tabla 9.2). Con los efectos activos se determinan las condiciones de operación más robustas, las cuales maximizan el valor de la respuesta S/R. Esto se hace mediante las gráficas de efectos, como se ilustrará más adelante.
2. Se realiza el análisis para la media $\bar{Y}$. Los factores que sólo afectan a la media se utilizan como factores de ajuste para llevar a ésta a su valor objetivo. Si todos los factores que afectan a la media también afectaran al estadístico S/R será necesario encontrar una solución de compromiso, utilizando como factor de ajuste el de más efecto en la media y con menos efecto en la dispersión o variabilidad.

Además de los pasos anteriores, para cualquier respuesta continua es útil realizar estos mismos dos pasos con el estadístico general $S/R = -10\ 10g(S^2)$, a fin de tener otra visión del análisis. Si los resultados de este otro análisis llegaran a diferir de los resultados del estadístico recomendado por Taguchi, se deben analizar con detenimiento las dos soluciones encontradas para finalmente determinar cuál es la mejor. Este segundo análisis puede proteger al experimentador de una solución subóptima que algunas veces generan las señales/ruido originales (Logothetis y Wynn, 1994).

Ejemplo 9.4

Análisis del experimento del color del pigmento. En el experimento de la figura 9.4 el color es una respuesta del tipo nominal es el mejor, con valor objetivo de 23. Por ello se decide utilizar el estadístico tipo II. Así, con los datos obtenidos de la figura 9.4 se aplican los dos pasos para la optimización.

1. Se analiza la variable $S/R = -10 \log(S^2)$ y se obtienen los efectos y sus gráficas, como se explicó en el capítulo 6. Las gráficas de efectos principales se muestran en la figura 9.8. Se recomienda utilizar el gráfico de Daniel para detectar efectos activos, como se hace en el diseño clásico (labor que se deja como ejercicio al lector). Aquí sólo presentamos las gráficas de efectos, y de ellas se observa que los efectos A y D son los que más afectan a la S/R. Es decir, los factores A y D influyen bastante sobre la variación del color del pigmento. De aquí que para maximizar la robustez se recomiende utilizar el factor A en su nivel alto y el factor D en su nivel bajo; este tratamiento hará más robusto al proceso (menos sensible al efecto de los factores de ruido que se han estudiado). Note que estos niveles ($A = 2$, $D = 1$) corresponden al renglón 5 y 7 del arreglo interno de la figura 9.4, y es en estas combinaciones de niveles donde se observa la menor variación; $S = 4.43$ y $S = 4.42$, respectivamente.
2. Se procede al análisis de la media, y las gráficas de efectos se muestran en la figura 9.9. Se observa que los factores B y D son los que tienen más efecto sobre la media. El factor B tiene efecto sobre la media, pero no tiene efecto sobre el estadístico S/R, de manera que puede utilizarse como factor de ajuste para llevar a la media a su valor nominal. Por ejemplo, si el valor objetivo del color es 23, conviene elegir el nivel alto del factor B. Por lo tanto, la solución propuesta es,

$$(A = 2, B = 2, C = \$, D = 1, E = \$, F = \$)$$

Aunque sólo se ha hecho énfasis en el uso de gráficas de efectos principales en los pasos anteriores, se debe hacer un análisis exhaustivo aplicando lo visto en capítulos anteriores, para lo cual es primordial el uso de un software. Por ejemplo en la tabla 9.3 se muestran los mejores ANOVA para la razón S/R y para la media. De donde se destaca que efectivamente los factores señalados antes son significativos. Además, los R^2 son altos.

De acuerdo con lo estudiado y con el ejemplo anterior, quedan claras las siguientes ventajas del diseño robusto con respecto del clásico:

1. *La introducción de la idea de robustez.* En el diseño clásico (capítulos 2 a 8) no se considera la posibilidad de minimizar el efecto de un factor de ruido, sin tener que controlarlo directamente, lo cual es la esencia del diseño robusto. Con el diseño robusto es posible formular un producto o proceso con las siguientes características: que trabaje sobre el valor objetivo con la mínima variación, que sea insensible a los cambios de las condiciones ambientales y que sea insensible a la variación transmitida por sus componentes y que al mismo tiempo tenga el menor costo de fabricación.

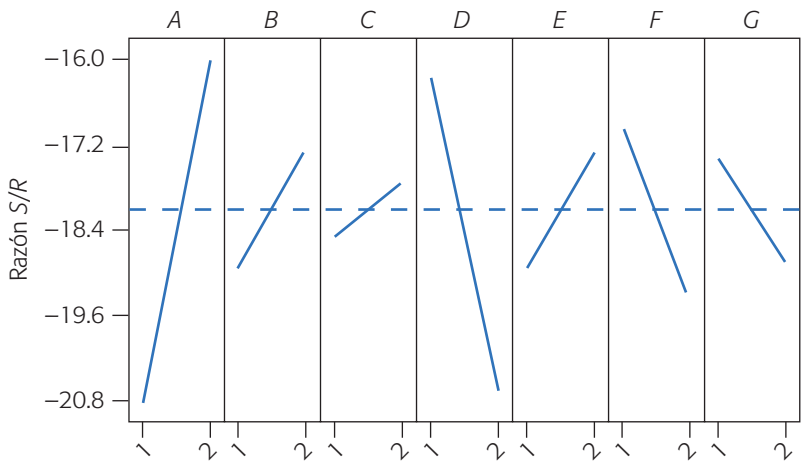


Figura 9.8 Gráficas de efectos para la variable señal/ruido (*S/R*), ejemplo del pigmento.

2. *Su énfasis en buscar siempre minimizar la varianza, y al mismo tiempo llevar la media a su valor nominal.* El diseño clásico centra su atención en mejorar el desempeño de la media del proceso, muchas veces ignorando el comportamiento de la varianza. En diseño robusto es típico enfocar el análisis al revés: primero se minimiza la varianza y después se lleva la media a su valor objetivo.
3. *El reconocer que es durante las etapas de diseño de un nuevo producto y en el diseño del proceso, cuando el diseño de experimentos puede tener mayor impacto.* De aquí el nombre alternativo de control de calidad fuera de línea (*off-line*) para las técnicas de calidad que se aplican en la etapa de diseño y

Tabla 9.3 Mejor ANOVA para razón *S/R* y la media. Ejemplo del pigmento.

Y: Razón *S/R:* $-10 \log S^2$.

Fuente	Suma de C.	*GI*	Cuadrado medio	Razón *F*	Valor-*p*
A	49.4	1	49.4	16.6	0.0103
D	37.6	1	37.6	12.23	0.0173
Error	15.4	5	3.08		
Total (corr.)	102.4	7			
$R^2 = 84.98\%$		R^2 (ajus) = 78.97%			

Y: Media.

Fuente	Suma de C.	*GI*	Cuadrado medio	Razón *F*	Valor-*p*
B	103.3	1	103.3	14.46	0.0126
D	118.2	1	118.2	16.54	0.0097
Error	35.7	5	7.15		
Total (corr.)	257.2	7			
$R^2 = 86.11\%$		R^2 (ajus) = 80.56%			

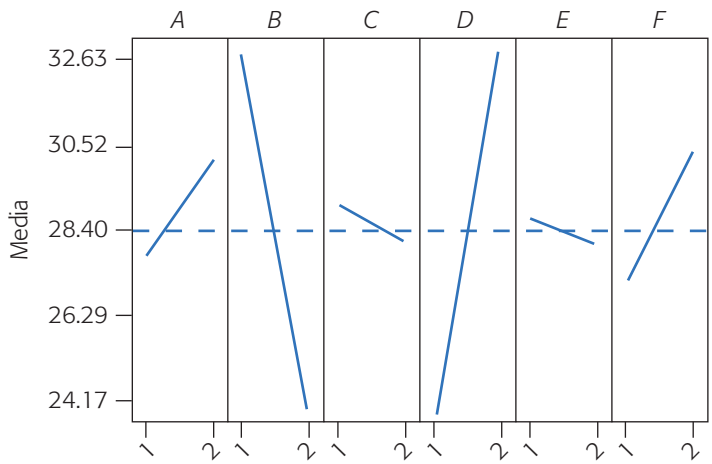


Figura 9.9 Gráfica de efectos para la media del color.

desarrollo de nuevos productos y procesos. Es precisamente en la etapa de diseño del producto cuando se pueden atacar las tres principales fuentes de variación que afectan su desempeño futuro, es decir, las variables ambientales, el deterioro del producto y las variaciones de manufactura. El diseño clásico no enfatiza lo suficiente el hecho de que una vez diseñado un producto y su correspondiente proceso, se reduce el margen de maniobra para optimizarlo, puesto que sólo queda por atacar la variación en manufactura.

4. *El diseño con arreglo interno y externo.* Este diseño experimental, introducido por Taguchi, es un mecanismo que permite introducir el efecto de los factores de ruido en cada combinación de los factores controlables. Por lo general, se supone que al menos para fines experimentales, los factores de ruido serán controlables, lo que permite manejarlos como parte del experimento.

Uso de software

Varios de los sistemas computacionales especializados en estadística incluyen diseños Taguchi, los cuales facilitan el trabajo de diseño y análisis. Por ejemplo, en *Statgraphics*, para acceder a estos diseños se sigue la secuencia: *Special* → *Experimental Design* → *Create Design* (o *Doe*), ahí se elige *Inner/Outer Arrays*, se señala el número de variables de respuesta así como el número de factores controlables y de ruido (*noise*) que se tienen en el experimento. Después de ello aparece una lista de posibles arreglos internos y externos que se pueden emplear en tal experimento y que son seleccionados de acuerdo con lo que se dijo antes. Después aparece una opción que permite asignar cada factor a las diferentes columnas de los arreglos que, como opción, ofrecen la recomendación que se proporciona en la figura 9.6. Al final se genera una tabla con los tratamientos que se correrán en el experimento. Para el análisis se debe elegir el tipo de razón *S/R*. En el caso de Minitab, la secuencia de opciones para los diseños robustos es *Stat* → *Doe* → *Taguchi* → *Create Taguchi Design*. En

donde se podrá indicar el número de factores para un arreglo ortogonal. En *Designs* también se encuentran los arreglos Taguchi que se pueden emplear.

Preguntas y ejercicios

1. Algunas de las diferencias básicas entre los diseños experimentales clásico y el robusto, se da en torno a los diferentes énfasis que cada uno hace con respecto a la media y a la varianza. Explique con detalle en qué consisten estas diferencias.
2. Comente algunas características de la función de pérdida de Taguchi y sus implicaciones.
3. ¿El diseño robusto propone controlar los factores de ruido o qué sugiere al respecto? Dé ejemplos.
4. Con base en los ejemplos vistos en el presente capítulo, proporcione un par de ejemplos de productos que ante ciertas circunstancias se desearía que fueran robustos.
5. Señale las cuatro categorías de un factor controlable con respecto a su efecto sobre la media y la variabilidad.
6. Explique qué es un factor señal y qué es un factor de ruido.
7. Los arreglos ortogonales son cierto tipo de diseños de experimentos, explique en qué consisten y proporcione un par de ejemplos.
8. ¿Para qué se recomienda emplear un arreglo L_{16}?
9. ¿En qué consiste un arreglo L_9 y cuándo se recomienda emplearlo?
10. Explique en qué consiste un diseño con arreglo interno y externo.
11. ¿Con qué propósito, la primera columna de un arreglo ortogonal tiene los niveles bastante agrupados y no aleatorizados? ¿Qué ventaja práctica tiene esto?
12. ¿Qué es una razón o cociente señal/ruido?
13. El análisis de un diseño robusto con arreglo interno y externo se hace en dos pasos. Explique con detalle estos pasos.
14. ¿Qué se sugiere cuando hay conflictos entre la solución del análisis del estadístico S/R y el análisis de la media?
15. ¿Cuáles son los inconvenientes que pueden surgir del análisis de los estadísticos S/R propuestos por Taguchi?
16. Diga al menos una desventaja que tiene el diseño con arreglo interno y externo.
17. ¿Para hacer diseño robusto es estrictamente necesario utilizar un diseño con arreglo interno y externo? Argumente.
18. Se conduce un experimento para encontrar una manera económica de ensamblar un conector elástico a un tubo de *nylon,* con la fuerza de unión requerida. Se tienen dos objetivos básicos: por un lado, minimizar el esfuerzo necesario para hacer el ensamble y, por el otro, maximizar la fuerza del ensamble realizado. Los investigadores identificaron cuatro factores de control y tres de ruido, que ellos pensaban podían estar afectando la fuerza del ensamble. Los factores y sus niveles fueron los siguientes:

Factores de control:

A. Interferencia (baja, media, alta).
B. Grosor de la pared (delgada, media, gruesa).
C. Profundidad de inserción (superficial, media, honda).
D. Porcentaje de adhesivo (bajo, medio, alto).

Factores de ruido:

M. Tiempo de acondicionamiento (24 h, 12 h).
N. Temperatura de acondicionamiento (72°F, 150°F).
O. Humedad relativa de acondicionamiento (25%, 75%).

Como se aprecia, se decide probar cada factor controlable en tres niveles y variar cada factor de ruido en dos niveles. Aunque no es posible controlar los factores de ruido durante la producción, sí hay manera de hacerlo para fines experimentales. Se seleccionó un arreglo ortogonal L_9 para los factores de control y un arreglo L_8 para los factores de ruido. El diseño con arreglo interno y externo resultante, así como los datos obtenidos, se muestran en la figura 9.10. Estos datos representan la fuerza necesaria para desunir el ensamble, el cual se busca maximizar.

a) De acuerdo con las recomendaciones de la figura 9.4, en el arreglo externo (L_8) se eligieron los renglones 1, 2 y 4 como los que definen los niveles de los factores *M,*

					Factores de ruido										
					M	2	2	2	2	1	1	1	1		
					N	2	2	1	1	2	2	1	1		
					MN	1	1	2	2	2	2	1	1		
					O	2	1	2	1	2	1	2	1		
					MO	1	2	1	2	2	1	2	1		
					NO	1	2	2	1	1	2	2	1		
					E	2	1	1	2	1	2	2	1	Media	Desviación estándar	Razón señal/ ruido
	Factores controlables															
	A	*B*	*C*	*D*										$\bar{X}$	*S*	
Arreglo interno	1	1	1	1	19.1	20.0	19.6	19.6	19.9	16.9	9.5	15.6				
	1	2	2	2	21.9	24.2	19.8	19.7	19.6	19.4	16.2	15.0				
	1	3	3	3	20.4	23.3	18.2	22.6	15.6	19.1	16.7	16.3				
	2	1	2	3	24.7	23.2	18.9	21.0	18.6	18.9	17.4	18.3				
	2	2	3	1	25.3	27.5	21.4	25.6	25.1	19.4	18.6	19.7				
	2	3	1	2	24.7	22.5	19.6	14.7	19.8	20.0	16.3	16.2				
	3	1	3	2	21.6	24.6	18.6	16.8	23.6	18.4	19.1	16.4				
	3	2	1	3	24.4	23.2	19.6	17.8	16.8	15.1	15.6	14.2				
	3	3	2	1	28.6	22.6	22.7	23.1	17.3	19.3	19.9	16.1				

Figura 9.10 Diseño con arreglos interno y externo del ejercicio 18.

N y *O*, respectivamente; para cada una de las ocho corridas. Tomando en cuenta esto, note que cada combinación de los factores de control se prueba en todas las combinaciones de los factores de ruido. Limítese a los renglones referidos antes y señale los niveles reales de estas combinaciones de los factores de ruido.

b) De acuerdo con lo que hizo en el inciso anterior, compruebe que las ocho condiciones o combinaciones de los factores de ruido conforman un diseño 2^3.

c) ¿Cuál es la razón señal/ruido adecuada para la fuerza del ensamble? Argumente.

d) Calcule la media, la desviación estándar y la razón señal/ruido en cada combinación de los factores de control, para cada combinación de los factores de control.

e) Realice el análisis de la razón señal/ruido, grafique los efectos principales de los factores de control y determine las condiciones de operación más robustas.

f) Haga el análisis para la media, grafique los efectos de los factores controlados y determine las mejores condiciones de operación para el proceso de ensamble.

g) Obtenga el mejor ANOVA para los dos casos anteriores.

19. En el proceso de fabricación de muelles para automóviles, una variable de salida importante es la altura sin carga, cuyo valor nominal es 8 pulgadas. La altura sin carga de un muelle se determina durante el tratamiento de calor en el cual se forma la curvatura del muelle. El problema es determinar las condiciones de este tratamiento que den por resultado la altura de carga deseada. Se corre un experimento con cuatro factores de control y uno de ruido. Los cuatro factores controlables son: (*B*) temperatura del horno, (*C*) tiempo de quemado, (*D*) tiempo de transferencia y (*E*) tiempo de permanencia (tiempo en el cual se forma la curvatura en una pieza caliente). Los ingenieros también estaban interesados en estudiar los efectos de interacción *BC, BD* y *CD*. El factor de ruido es la temperatura del lubricante (*O*), que es difícil de controlar durante la producción. Los niveles utilizados en cada factor se muestran en la siguiente tabla:

Etiqueta	Factor	Niveles	
		Bajo	Alto
B	Temperatura del horno (°F)	1 840	1 880
C	Tiempo de quemado (seg.)	25	23
D	Tiempo de transferencia (seg.)	12	10
E	Tiempo de permanencia (seg.)	2	3
O	Temperatura de lubricante (°F)	130-150	150-170

Note que, debido a lo difícil que es controlar de manera puntual los niveles del factor de ruido, se utilizan rangos de temperatura. Se emplea un arreglo ortogonal L_8 para los factores controlables. Los datos obtenidos se muestran en la tabla 9.4.

a) Analice la matriz de diseño para los factores de control y diga a qué diseño factorial fraccionado corresponde.

b) ¿Qué resolución tiene tal fracción factorial? Escriba su estructura alias reducida.

c) ¿Cuál es el estadístico *S/R* más apropiado al tipo de variable de respuesta que se tiene? Argumente.

d) Calcule el estadístico *S/R* tipo I, analice con detalle la respuesta *S/R* y determine el tratamiento más robusto.

e) ¿Tienen efecto sobre la razón señal/ruido las interacciones que les interesaban a los ingenieros?

f) Analice la respuesta promedio $\bar{Y}$.

Tabla 9.4 Datos del ejercicio 19.

							Factor de ruido					
							O^-			O^+		
B	*C*	*BC*	*D*	*BD*	*CD*	*E*						
−	−	+	−	+	+	−	7.78	7.78	7.81	7.50	7.25	7.12
+	−	−	−	−	+	+	8.15	8.18	7.88	7.88	7.88	7.44
−	+	−	−	+	−	+	7.50	7.56	7.50	7.50	7.56	7.50
+	+	+	−	−	−	−	7.59	7.56	7.75	7.63	7.75	7.56
−	−	+	+	−	−	+	7.94	8.00	7.88	7.32	7.44	7.44
+	−	−	+	+	−	−	7.69	8.09	8.06	7.56	7.69	7.62
−	+	−	+	−	+	−	7.56	7.62	7.44	7.18	7.18	7.25
+	+	+	+	+	+	+	7.56	7.81	7.69	7.81	7.50	7.59

g) Si hay algún factor que no afectó al cociente S/R, pero que sí afecta la media, utilícelo como factor de ajuste para acercar la media a su valor objetivo.

h) Repita el análisis utilizando el estadístico $S/R = -10 \log(S^2)$ y comente las diferencias observadas.

20. En 1951 en la Ina Tile Company se tenía el problema de que el horno quemaba de forma dispareja debido a una variación de la temperatura en diferentes partes de éste, lo cual causaba defectos en las lozas que se fabricaban. Una posibilidad de solución (imposible en ese momento) era cambiar el horno por otro que no tuviera ese problema. Otra posibilidad era reformular las lozas de manera que fueran robustas al funcionamiento "disparejo" del horno. Esto último fue lo que se decidió hacer, utilizando los siguientes niveles de prueba en siete factores de la formulación de la loza:

Factor	Nivel 1	Nivel 2
A: Aditivo de cal	$A_1 = 5\%$	$A_2 = 1\%$ (actual)
B: Granularidad del aditivo	$B_1 =$ tosca (actual)	$B_2 =$ fina
C: Contenido de algamatolite	$C_1 = 43\%$	$C_2 = 53\%$ (actual)
D: Tipo de algamatolite	$D_1 =$ mezcla actual	$D_2 =$ más barata
E: Cantidad de carga	$E_1 = 1\,300$ kg	$E_2 = 1\,200$ kg (actual
F: Contenido de reciclado	$F_1 = 0\%$	$F_2 = 4\%$ (actual)
G: Contenido de feldespato	$G_1 = 0\%$	$G_2 = 5\%$ (actual)

Note que uno de los niveles de prueba para cada uno de los factores corresponde al nivel que se utilizaba hasta ese momento. Se tomó una muestra de 100 lozas en cada uno de los ocho tratamientos y se obtuvo el porcentaje de lozas defectuosas. Los resultados obtenidos se muestran en la siguiente tabla:

Núm. de corrida	*A*	*B*	*C*	*D*	*E*	*F*	*G*	% de lozas defectuosas
1	1	1	1	1	1	1	1	16
2	1	1	1	2	2	2	2	17
3	1	2	2	1	1	1	2	12
4	1	2	2	2	2	2	1	6
5	2	1	2	1	2	2	2	6
6	2	1	2	2	1	1	1	68
7	2	2	1	1	2	2	1	42
8	2	2	1	2	1	1	2	26

a) ¿Por qué este experimento es un diseño robusto?
b) Analice con detalle los datos: efectos principales y efectos activos.
c) Obtenga la mejor formulación de las lozas. Asigne el nivel más económico a los factores que no tienen efecto sobre el porcentaje de defectuosos.
d) ¿Cuál es la proporción de loza defectuosa esperada en el tratamiento elegido?
e) Estime la diferencia entre la proporción de loza esperada en el tratamiento anterior (actual) y el tratamiento nuevo sugerido por el estudio.

21. Con respecto al ejemplo 9.4, haga un análisis más exhaustivo: gráfica de Pareto, gráfica de Daniel y obtenga el mejor ANOVA. Con base en esto y en las gráficas de efectos que ya se mostraron en el ejemplo, confirme las conclusiones obtenidas.

22. Analice con detalle los datos experimentales del diseño robusto del ejemplo 14.4 (harina robusta) del capítulo 14, y contraste las conclusiones principales que obtenga con las que se consiguieron en el capítulo 14, cuando este diseño se analizó con la técnica de parcelas divididas.

Capítulo 10
Planeación de un experimento

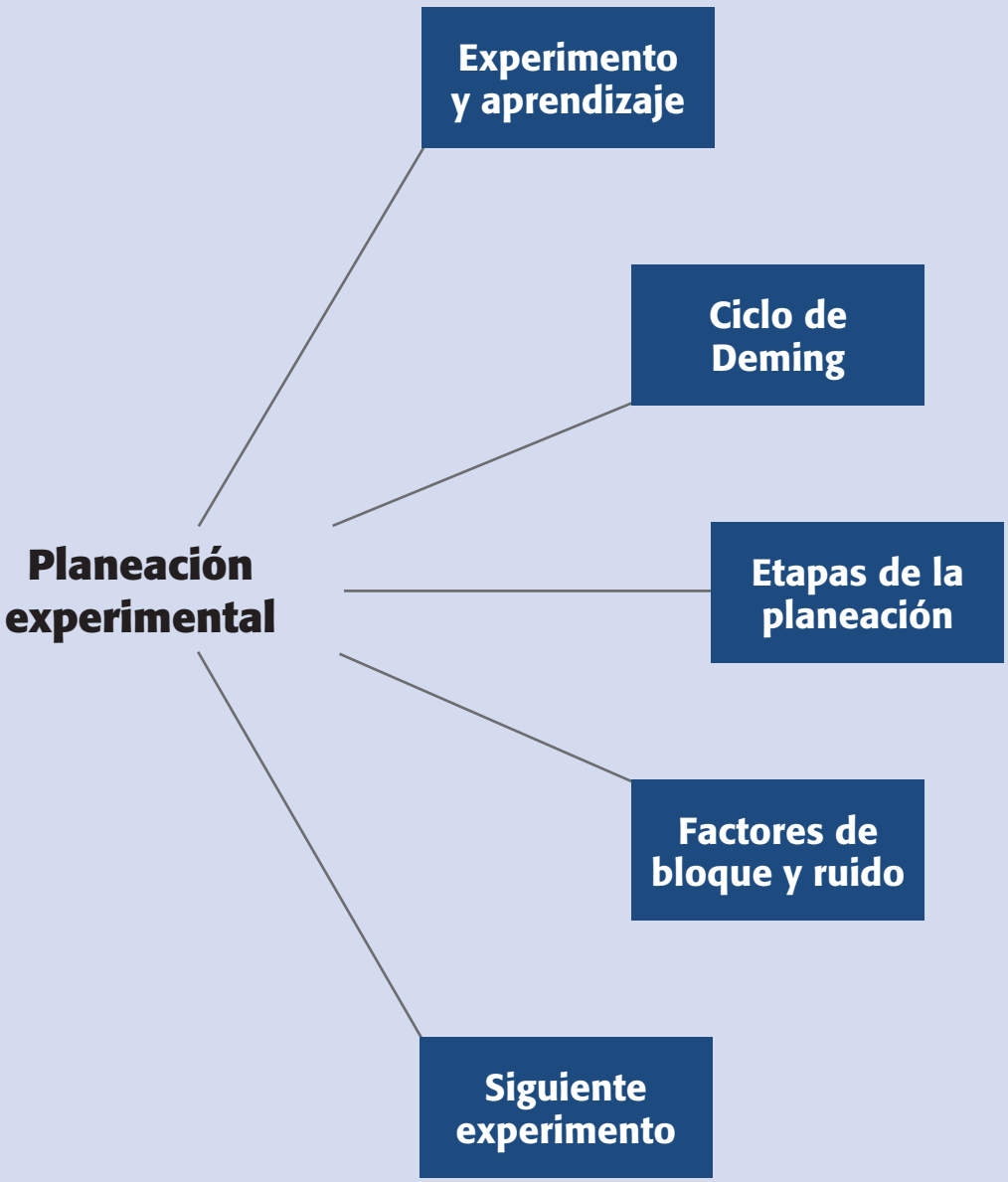
Sumario

- Experimentación: una estrategia para probar conjeturas y generar aprendizaje
- El diseño de experimentos y el ciclo de Deming
- Etapas y actividades de la planeación y análisis de un experimento
- Control de factores de bloque y de ruido
- Qué sigue después del primer experimento
- Qué hacer cuando ningún efecto es significativo

Objetivos de aprendizaje

- Explicar la importancia de la experimentación como estrategia para generar aprendizaje sobre un proceso.
- Describir el ciclo de Deming y cada una de las etapas de la planeación de experimentos.
- Aplicar el control a factores de bloque y de ruido como elementos de la planeación de los experimentos.
- Entender las alternativas a seguir después de un primer experimento.

Mapa conceptual



Conceptos clave

- Ciclo de Deming
- Factores
- Modelo de primer orden
- Nivel variable de un factor
- Variables de salida

Experimentación: una estrategia para probar conjeturas y generar aprendizaje

Al inicio del capítulo 1 señalamos que, en la industria, es una práctica común hacer experimentos o pruebas con la idea de que al mover o hacer algunos cambios sobre los materiales, métodos o condiciones de operación de un proceso se puedan detectar, resolver o minimizar los problemas de calidad. También comentamos que es común que estas pruebas o experimentos se hagan sobre la marcha, a prueba y error, con base en la experiencia y la intuición, en lugar de seguir un plan experimental adecuado que garantice una buena respuesta a las interrogantes planteadas. De aquí que, más que apostarle a la improvisación y a la estrategia de prueba y error, el reto es mejorar la forma en que se diseña un experimento.

De hecho, el éxito de un experimento radica en el alto porcentaje en la calidad de su planeación. Por ello, es importante contener las ansias y no hacer pruebas antes de conceptualizar el problema y decidir la mejor estrategia experimental para abordarlo. En los ejemplos y ejercicios de los capítulos anteriores, hay evidencia abundante del poder del diseño de experimentos; sin embargo, para que este poder se traduzca en resultados es necesaria una buena planeación.

El diseño de experimentos se puede aplicar a problemas o situaciones en las que se quiere investigar y/o probar conjeturas en las que se comparan dos o más situaciones para las causas o factores involucrados.

Ejemplo 10.1

En una industria electrónica hay un proceso llamado Soldadora de ola, en el que se tienen problemas por excesos e insuficiencias de la soldadura en tarjetas electrónicas. En el proceso hay muchos factores controlados que pueden estar causando el problema, por ejemplo la temperatura del flux (sustancia que provee de adherencia a la soldadura), la temperatura de la soldadura, altura la de la ola, la velocidad del proceso, etc. Naturalmente que aquí se podría aplicar diseño de experimentos para encontrar las condiciones de estos factores controlados que minimicen el problema. Se podría empezar con un diseño factorial (quizá fraccionado) para encontrar los factores que mayor influencia tienen.

Sin embargo, en una situación particular, más que sospechar de tales factores controlados, los responsables del proceso consideraban otra situación. Tienen la conjetura de que la rejilla por donde se expulsa el flux se tapa en algunos orificios con relativa frecuencia, y esto puede deberse a una o más de las siguientes causas: al desprendimiento de fibras de los guantes de tela que utilizan los operarios para el acomodo de las tarjetas en los *pallets* antes de que pasen a la soldadora; a la acumulación de residuos en los *pallets* que se desprenden cuando son bañados por el flux, y al utilizar demasiado tiempo el mismo flux (flux sucio). Saben que si se tapa la rejilla se genera el problema, pero no hay una forma fácil de verificar que en realidad esto sucede, por lo que es necesario investigar las situaciones que supuestamente favorecen el bloqueo de la rejilla.

Note, en este último caso, que no se sospecha de factores controlados, sino de situaciones muy particulares del proceso. Lo más "normal" en estos casos sería tomar medidas para atender esas supuestas causas y esperar a ver si eso disminuye el proble-

ma; sin embargo, esto no es lo más recomendable, ya que en ocasiones puede causar incomodidades adicionales a los operarios y mayores costos de operación del proceso. Por ello, lo ideal sería verificar o confirmar las conjeturas antes de tomar decisiones.

En este tipo de casos, donde las causas no son factores controlados, para aplicar el diseño de experimentos se debe idear una forma de contrastar la condición de riesgo contra una de menor riesgo. Por ejemplo, en este caso se decidió correr un diseño factorial 2^4 utilizando los siguientes factores y niveles: tipo de guante usado por el operador (el usual de tela o de látex); estado del *pallet* (con ocho horas de uso o recién lavado); frecuencia con la que se cepilla la rejilla del flux (no cepillarla durante la corrida o cepillarla cada media hora), y estado del flux (flux con 24 horas de uso aproximadamente o flux limpio). Cada corrida experimental consistió en soldar 50 tarjetas y contar el número de insuficiencias y/o excesos encontrados en ellas.

Es difícil cambiar el factor de flux de una corrida a otra, ya que si se tiene flux limpio y en la siguiente corrida toca flux sucio, habrá que esperar 24 horas aproximadamente para que el flux esté realmente sucio. Es decir, en la práctica se tiene un problema de aleatorización para uno de los factores. En estos casos se puede aplicar un diseño en parcelas divididas (véase capítulo 14) o apelar a las enseñanzas de Taguchi (véase capítulo 9), y sacrificar parcialmente la aleatorización con tal de efectuar el experimento. En el aspecto operativo esto se resolvió de la siguiente manera:

- Se generó de manera aleatoria una matriz de diseño para un diseño 2^4. Esta matriz se muestra en la primera parte de la tabla 10.1.
- Con el análisis de la tabla 10.1 se aprecia que la columna del factor *C* es la que está menos "aleatorizada", es decir, es la que tiene más rachas con signo igual de manera consecutiva. Por ejemplo, hasta las primeras siete corridas sólo hay un signo +. Por esta razón se decidió asignar al estado el flux como

Tabla 10.1 Matriz de diseño para el ejemplo 10.1.

Orden original				
Corrida	*A*	*B*	*C*	*D*
1	1	−1	−1	−1
2	1	−1	−1	1
3	−1	1	1	−1
4	1	1	−1	1
5	−1	−1	−1	−1
6	−1	−1	−1	1
7	−1	1	−1	−1
8	1	1	1	−1
9	−1	−1	1	1
10	1	1	1	1
11	−1	1	−1	1
12	1	−1	1	1
13	1	1	−1	−1
14	−1	−1	1	−1
15	1	−1	1	−1
16	−1	1	1	1

⇒

Orden modificado por factor *C*				
Corrida	*A*	*B*	*C*	*D*
1	−1	−1	−1	−1
2	1	−1	−1	1
4	1	1	−1	1
5	−1	−1	−1	−1
6	−1	−1	−1	1
7	−1	1	−1	−1
3	−1	1	1	−1
8	1	1	1	−1
9	−1	−1	1	1
10	1	1	1	1
11	−1	1	−1	1
13	1	1	−1	−1
12	1	−1	1	1
14	−1	−1	1	−1
15	1	−1	1	−1
16	−1	1	1	1

el factor *C*, e iniciar el experimento con flux sucio. A la corrida 3 se le cambió el orden y de esta forma las primeras seis corridas implican el factor flux en –1; adicionalmente, la corrida 12 se intercambió por la 13 (véase tabla 10.1). Si bien es cierto que con esto se sacrifica la aleatorización de uno de los factores que podría confundirse con los posibles efectos temporales, es mejor que no estudiar tal factor. De hecho, los arreglos Taguchi de entrada presuponen que hay factores difíciles de aleatorizar, por lo que proponen una asignación de los factores de acuerdo a esta dificultad.

La idea del ejemplo anterior es ilustrar que, cuando se sospecha de diferentes causas para un problema, se puede idear una forma de contrastar o verificar cada una de esas causas, y cuando eso se logra, con un solo estudio experimental es posible investigar todas las causas en forma conjunta. Además, cuando hay problemas de aleatorización se pueden buscar soluciones intermedias, como la que se mostró, para así efectuar el diseño factorial. Un buen ejemplo de un diseño experimental en donde se ven situaciones "difíciles" de estudiar, es el experimento que hizo Ossie Cadenza en su bar (ejercicios 15 y 16 del capítulo 8). Así, cuando hay conocimiento del diseño de experimentos, se conoce el proceso y se aplica un poco de imaginación, se puede diseñar un experimento para investigar diferentes conjeturas sobre una situación o problema.

La dirección del aprendizaje es entonces: conocimiento, conjeturas, imaginación, experimentación, análisis y reflexión; y de forma repetida, esto se convierte en aprendizaje. Para establecer conjeturas no se trata de considerar en el experimento una lista arbitraria de factores o situaciones de los cuales no se sabe nada; en cambio, se debe analizar la información disponible sobre el proceso (datos históricos, experimentos previos, estudio de caracterización) para, en la medida de lo posible, hacer conjeturas más concretas de la relación entre la variable de respuesta y los factores a investigar, ya que esto permitirá planear mejor el experimento que se necesita. Si después de reflexionar y analizar la situación no se llega a ninguna idea útil, entonces, al estudiar cómo afectan los factores controlables a la respuesta es válido plantear el experimento sólo como un medio para ganar conocimiento sobre el proceso.

A menudo sucede que el primer diseño experimental da lugar a que se continúe experimentando en cierta dirección, ya que se generan nuevos conocimientos e ideas. La *experimentación* es un proceso interactivo, como el aprendizaje. Este proceso interactivo se ilustra en las figuras 1.1 y 1.2.

El diseño de experimentos y el ciclo de Deming

Ciclo de Deming
Estrategia que busca mejorar continuamente un proceso, mediante la aplicación repetida del ciclo: planear, hacer, verificar y actuar.

Muchas estrategias de optimización y de mejora continua, y en particular el diseño de experimentos, embonan en el llamado *ciclo de Deming*, el cual representa una estrategia a seguir para mejorar continuamente cada producto o proceso, y consiste en los cuatro pasos: *planear*, *hacer*, *verificar* o *estudiar y actuar*, que se representan en la figura 10.1. A continuación se describe a detalle este ciclo en el contexto de un diseño de experimentos.


4. Actuar de acuerdo con los resultados: implemente el cambio o deséchelo, o bien, planee una prueba.
1. Planear un cambio o una prueba con el objetivo de encontrar causas y lograr una mejora.
Ciclo de Deming
2. Hacer o llevar a cabo el cambio o las pruebas planeadas (de preferencia a pequeña escala).
3. Estudiar los resultados. ¿Qué se aprendió? ¿Qué estuvo mal?

Figura 10.1 Ciclo de Deming y los experimentos.

1. *Planear un experimento*. Lo primero es localizar un problema importante que sea susceptible de atacarse con diseño de experimentos. La importancia del problema se puede cuantificar con base en la pérdida que causa a la compañía. Una vez ubicado el problema se determinan los factores controlables que pudieran tener alguna influencia sobre la característica de calidad de interés. Se diseña el plan experimental más adecuado a la situación, con lo cual quedan especificadas las corridas del proceso que han de realizarse: cuáles tratamientos y cuántas repeticiones de cada uno.
2. *Hacer las corridas experimentales*. Se prueba el proceso en cada tratamiento especificado en la etapa de planeación. Las corridas se hacen en orden aleatorio y de acuerdo al plan experimental seleccionado.
3. *Verificar o analizar los resultados*. Consiste en analizar los datos con las técnicas adecuadas. Verificar los supuestos y determinar el mejor tratamiento (o tratamiento ganador) para la variable de respuesta que es objeto de estudio.
4. *Implementación de los resultados*. Hacer pruebas confirmatorias para tener una mejor representatividad del desempeño del proceso en el nuevo tratamiento. Con base en lo aprendido es posible planear otro experimento y repetir el ciclo de Deming. Cada vuelta al ciclo implica necesariamente mejoras, ya que en el peor de los casos, cuando no se consigue la mejora se ha logrado más conocimiento del proceso. A continuación veremos con mayor detalle el ciclo de Deming aplicado al diseño de experimentos.

Etapas y actividades de la planeación y análisis de un experimento

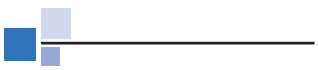
En el capítulo 1 se presentaron en forma breve las etapas a cumplir en el diseño y análisis de experimentos, aquí se vuelven a presentar, señalando algunas actividades específicas que se deben desarrollar.

Planeación y diseño

1. *Definir el problema o el objetivo.* En este primer paso se debe hacer un esfuerzo especial por entender el problema o el objetivo que se busca al experimentar, ya que aquí se sentarán las bases para que el experimento sea exitoso. Por ello se recomienda desarrollar las siguientes actividades:

   a) Asignar un título al estudio experimental que indique la intención general del mismo.
   b) Describir el problema o situación destacando los argumentos de por qué es importante hacer el estudio. Se recomienda utilizar cifras que indiquen la importancia de la situación (defectos, quejas, oportunidades, etcétera).
   c) Especificar la manera en que se mediría el éxito del proyecto experimental. De preferencia definir uno o varios indicadores que ayuden a ello.
   d) Comentar qué se hace actualmente para atenuar el problema o si se cuenta con experiencias de experimentos anteriores.
   e) Definir el objetivo que se persigue al realizar el experimento, en donde quede claro el resultado deseado.

2. *Hacer un esquema del estudio donde se señale el problema planteado*. Destacar las variables de salida (respuestas), las variables de operación (parámetros) del proceso y las principales entradas (materiales, sustancias, etcétera).



Factores
Son las causas o variables a estudiar en un experimento, debido a su posible impacto en la variable de respuesta de interés.

3. *Determinar los factores que deben investigarse, de acuerdo a su posible impacto en el problema.* Para seleccionar de manera adecuada los factores (o causas) a investigar se recomienda desarrollar las siguientes actividades:

   a) Hacer una lista de todas las posibles causas o variables independientes involucradas en el problema. Se recomienda que en la etapa inicial de una investigación experimental se consideren las causas sobre las que se tiene una seguridad razonable, es decir, no es necesario estar seguro que una causa es importante para considerarla en la investigación experimental.
   b) Para cada una de las causas principales del punto anterior (que llamaremos *factores*), señalar cómo se corroborarían, con una prueba experimental, que efectivamente son una causa y/o solución importante. Utilice la imaginación, el conocimiento y la experiencia.
   c) Decidir sobre los factores a estudiar. La recomendación es incluir en el experimento a cada factor que pueda tener una influencia importante en la respuesta. Cuando de entrada son muchos los factores (más de 10), algunos se podrán descartar con un análisis del historial del proceso, usando toda la información disponible y los conocimientos técnicos del proceso. Si después de esto el número de factores aún es grande, en primera instancia se puede correr un diseño factorial fraccionado saturado que permita detectar con un mínimo de pruebas aquellos factores que parecen tener mayor influencia en la respuesta (véase capítulo 8). En una segunda etapa experimental se pueden realizar experimentos más informativos sobre los factores que resultan importantes.

d) Elegir los niveles de prueba para cada factor. Para probar el impacto de un factor en general bastan dos niveles, salvo que por su naturaleza incluya tres o más niveles. También es importante la separación que se elija entre un nivel y otro en un factor dado. Los niveles muy juntos no sirven para detectar el efecto de ese factor, mientras que niveles muy separados pueden causar problemas en la operación del proceso. Se debe recurrir al conocimiento de la física del proceso, de tal forma que los niveles representen opciones de búsqueda sobre en qué nivel se debe operar el proceso.

En ocasiones, las costumbres en la operación del proceso limitan en un rango muy estrecho los posibles niveles de un factor. Sobre esto, es importante tener espíritu de investigador y así como cierto nivel de "incredulidad" sobre tales costumbres, para tratar de seleccionar niveles de prueba que realmente sean diferentes. En un experimento se buscan soluciones y también generar conocimiento.

Hay situaciones en donde es difícil fijar en forma más o menos exacta el nivel de un factor, tanto durante el experimento como en condiciones reales; entonces, en este caso quizás el *nivel de un factor* sea un rango de variación, cuyo centro es el nivel "teórico" de tal factor, pero en este caso se debe tener cuidado de que los rangos en los que se movería el factor en cada nivel no se traslapen. De esta forma, hay que asegurarse de que para cada factor se logre tener por lo menos dos condiciones de prueba realmente diferentes.

Nivel variable de un factor
Cuando no se puede fijar el nivel de un factor en un valor específico, se define dicho nivel como un pequeño rango de operación.

Si el factor no es numérico, se deben seleccionar por lo menos dos situaciones representativas de tal factor. Por ejemplo, si el factor es experiencia de la mano de obra, entonces se deben seleccionar como niveles de tal factor a dos operadores con experiencia muy diferente. Pero si el factor es la influencia de una operación previa en el material, entonces los niveles pueden ser hacer y omitir tal operación. En ocasiones será necesario utilizar más de dos niveles debido a una fuerte sospecha de curvatura. También puede ocurrir que, por su naturaleza, el factor tenga un número finito (digamos cuatro) de niveles, en cuyo caso es de interés para el experimentador probarlos todos en el estudio.

Cuando no se sabe cuáles niveles utilizar en un factor, es recomendable realizar corridas preliminares[1] del proceso, moviendo sólo ese factor y con esta información seleccionar los niveles a utilizar en el experimento.

[1] Se eligen dos niveles del factor y se realizan un par de corridas del proceso en cada nivel, manteniendo fijos todos los factores restantes. Si los resultados promedio en cada nivel quedan separados (una distancia de al menos 1.5 sigmas del proceso), es que ya se encontraron dos niveles adecuados para el factor. Pero si la distancia es menor, se abren un poco más los niveles y se vuelven a obtener datos, y así hasta separar las medias a la distancia requerida. Si el factor tiene poco o nulo efecto, será prácticamente imposible encontrar dos niveles que satisfagan la condición, en cuyo caso se eligen los niveles más abiertos. No se debe omitir este factor, puesto que su efecto importante se puede manifestar al interactuar con los demás factores en el experimento.

Variables de salida
Característica de calidad o desempeño que se quiere mejorar con el experimento.

4. *Elegir la(s) variable(s) de respuesta que será(n) medida(s) en cada punto del diseño y verificar que se miden de manera confiable.* Estas variables son el objetivo del experimento, ya que son las *variables de salida* o críticos de la calidad que se quieren corregir o impactar. Para encontrar estas variables realice las siguientes actividades:

   a) Haga una lista de las variables de salida o características de calidad del producto en donde se espera que se reflejen los cambios en los factores controlados que se estudiarán.
   b) De la lista anterior, seleccione una o varias, tomando en cuenta la facilidad con la que se pueden medir, el mayor impacto que tendrían sobre ellas los factores a estudiar y que reflejen mejor la magnitud del problema. Es importante considerar como variables de respuesta, además de las que están relacionadas con el problema, aquellas que pueden resultar afectadas por los cambios en los factores, aunque actualmente no sean un problema. Por ejemplo, si el tiempo de producción no es problema, pero con alguno de los factores se podría afectar éste, entonces sería bueno medirlo en las pruebas experimentales.
   c) Para las variables de respuestas seleccionadas, es preciso asegurarse de que se pueden medir de manera confiable. En otras palabras, revisar la forma en que se mediría cada variable de respuesta, y ver si el proceso de medición es confiable (equipos, gente, métodos). Por ejemplo, si es una variable que ordinariamente se mide, entonces es necesario investigar si ya existe un estudio R&R que respalde el proceso de medición (Gutiérrez Pulido y De la Vara, 2004). De no ser así, habrá que pensar en realizar uno.

5. *Seleccionar el diseño experimental adecuado a los factores que se tienen y al objetivo del experimento*. Aquí se debe elegir el conjunto de pruebas que se van a correr, y esto resulta de la cantidad de factores y niveles seleccionados; también considerar el costo y tiempo necesarios. Por ello, es preciso desarrollar las siguientes actividades:

   a) Con base en los factores y niveles seleccionados determinar si es factible aplicar un diseño factorial. Es decir, verificar si es posible correr en el proceso cualquier combinación de niveles de factores o si hay situaciones como las siguientes:

      - Por alguna razón se dificulta la aleatorización de un factor, ya sea porque cambiar de un nivel a otro es tardado, difícil o costoso. En este caso se pueden buscar soluciones similares a las que se presentaron en el factor flux del ejemplo 10.1, o si el factor no se considera muy decisivo, quizá se pueda estudiar como bloque.
      - Algunos tratamientos son difíciles o imposibles de correr, como la combinación con los niveles altos de todos los factores. En este caso si se tienen muchos factores, entonces el diseño que se aplicará probablemente sea un factorial fraccionado y, por lo tanto, se seleccionaría la

fracción que no tiene esos tratamientos difíciles de correr. En caso de que no se llegue a un factorial fraccionado, habrá que reconsiderar la selección de niveles para el factor que impide en mayor medida correr el o los tratamientos problema.

b) A partir de lo anterior y pensando en las variables de respuesta que se van a medir, es necesario decidir en qué consistiría cada prueba; por ejemplo, si se está experimentando con un proceso masivo y rápido, será importante determinar el tamaño de prueba que defina cuánto tiempo se va a dejar operando el proceso en cada condición experimental. En estos casos, cuando la variable de respuesta es un conteo, en la ecuación 6.13 de la sección "Experimento 2^3: ejemplo integrador" del capítulo 6 se da una fórmula para determinar el tamaño de la prueba en función de la magnitud del problema, y en los ejercicios 9, 14 y 17 del capítulo 6 se muestran casos en los que se determinó el tamaño de la prueba. Por otro lado, cuando se tienen procesos rápidos y masivos pero la variable de respuesta es de tipo continuo, generalmente se procesan varios productos (de 5 a 20) en cada condición experimental, y se miden todos o una muestra de los productos. Sin embargo, en lugar de analizar los datos individuales se analiza su media y su desviación estándar, véase el ejercicio 11 del capítulo 6 para mayores detalles al respecto.

c) Proponer un primer diseño que tome en cuenta el número total de corridas, costos, tiempos, etcétera.

d) Investigar si hay algún factor de ruido o bloque que podría estar actuando durante el experimento, por ejemplo, si el material de prueba que se utilizaría es suficientemente homogéneo o hay diferencia de lote a lote, en cuyo caso será necesario bloquear el efecto de lote. Por ejemplo, si se tiene en mente correr un diseño factorial 2^k y es necesario correrlo en diferentes bloques, entonces se debe decidir qué corridas deben ejecutarse en cada bloque (se sugiere revisar el capítulo 6 y/o apoyarse en el software adecuado).

e) Decidir el diseño específico que se correrá incluyendo el orden (aleatorio) en el que se efectuarán las pruebas. Es preciso verificar que: con el diseño elegido se puede cumplir el objetivo, es posible superar de manera adecuada las situaciones o restricciones prácticas y el diseño se puede correr con un costo y tiempo razonables. Para el número de réplicas, se deben seguir las recomendaciones que se han dado para cada diseño en particular, véase por ejemplo la sección "El tamaño de la muestra" del capítulo 3, y para el diseño 2^k consulte la tabla 6.12 del capítulo 6.

Recuerde que cuando sólo interesan los efectos lineales o principales de los factores, es suficiente un diseño factorial fraccionado de resolución III, que además es el más económico. Si también interesan algunas o todas las interacciones entre ellos, será necesario utilizar un diseño factorial fraccionado de resoluciones IV o V. Pero, si se esperan efectos de curvatura pura de algunos o todos los factores, entonces se deberá incluir un punto al centro del diseño para detectar la presencia de curvatura, y si son pocos factores, se

espera que el diseño seleccionado de una vez permita estudiar los efectos de curvatura pura.

Un aspecto crítico a la hora de seleccionar un experimento son los costos que implicará llevarlo a cabo, ya que se deben considerar todos los recursos que se gastarán en el experimento, entre los que se encuentran: material, energía, tiempo hombre y tiempo máquina. Debe buscarse que el gasto de recursos sea lo menor posible, al mismo tiempo que se cumplen los objetivos del experimento con la precisión deseada.

6. *Planear y organizar el trabajo experimental.* A partir del diseño seleccionado, organizar y planear con detalle el trabajo experimental. Algunos detalles a contemplar son los siguientes:

   a) Diseñar una hoja de trabajo en la que se especifique de manera clara cada prueba, el orden en que será corrida y la forma en que se medirán los resultados o se colectarán los datos de la prueba. Por lo general, los programas de software especializados proporcionan un formato u hoja de trabajo para este propósito.
   b) Definir a los responsables del proyecto, a las personas que van a efectuar los experimentos, así como las instrucciones generales que habrá que darles.
   c) Detallar instrucciones específicas que se seguirán en ciertas pruebas experimentales. Por ejemplo, si se experimenta con factores lentos, como la temperatura en un proceso, antes de correr el experimento o de empezar a registrar los resultados del proceso, será necesario esperar a que las condiciones con las que se quiere que opere el proceso efectivamente se logren.
   d) Completar todos los detalles de logística, como días, hora, materiales que se utilizarán y máquinas donde se harán las pruebas.
   e) Prever algunas posibles contingencias que podrían ocurrir durante el experimento, así como las acciones a realizar en caso de que ocurrieran.
   f) Si el costo no es muy elevado, hacer una prueba de ensayo con cualquier tratamiento, en el cual participen todos aquellos que están planeando el experimento, con miras a afinar imprevistos y detallar más las instrucciones.

7. *Realizar el experimento*. Aplicar el plan previsto en la etapa previa, y en caso de algún imprevisto no contemplado, reportarlo a los responsables para que se decida qué hacer.

Análisis e interpretación

8. *Hacer un análisis detallado de los resultados experimentales.* Antes de aplicar cualquier análisis se debe echar un vistazo a los datos para verificar que no hay errores obvios en su registro, así como detectar a simple vista las principales tendencias. A lo largo de los capítulos hemos hecho énfasis en los pasos a seguir para analizar cada uno de los diseños de experimentos

presentados. El análisis estadístico y el análisis de varianza dependen del diseño empleado. Se recomienda apoyarse en gráficas y pruebas analíticas.

9. *Interpretar resultados.* Una vez que se ha depurado el modelo, y se tiene el mejor ANOVA, es preciso analizar con detalle lo que ha pasado en el experimento, para ello se contemplan los siguientes aspectos:

   a) Ver los factores y efectos que influyeron de manera significativa en las diferentes variables de respuesta, y también señalar los factores que no tuvieron un impacto considerable. Contrastar esto con las conjeturas previas y señalar lo más importante.
   b) Interpretar con detalle los efectos más significativos para cada variable de respuesta, considerar, por ejemplo, la manera en que la variable de respuesta responde a los cambios en los diferentes factores. Es preciso hacer énfasis en las interacciones, y estudiar cómo estos efectos ayudan a conocer mejor el proceso y a corroborar o desechar conjeturas, desde la óptica de la física del proceso.
   c) Encontrar el tratamiento ganador, es decir, en qué condiciones se propone operar el proceso. Considerar las diferentes variables de respuesta y también los aspectos económicos. Si el experimento tuvo varias respuestas de igual importancia y en todas se tiene un modelo estadístico que las describe, entonces será necesario aplicar una estrategia de optimización multirrespuesta (véase capítulo 13).
   d) Verificar los supuestos del modelo, que por lo general son *normalidad*, *varianza constante* e *independencia de los residuos*. Si los supuestos se violan gravemente, ver en qué sentido afectan a las conclusiones. Si éstas resultan muy afectadas, entonces ver por qué los supuestos no se cumplen y actuar en consecuencia. Por ejemplo, si el problema es la falta de varianza constante, pero el tratamiento ganador se ubica entre los que tienen menor varianza, entonces no hay mayor problema. Si está entre los que tienen mayor varianza, entonces habrá que transformar los datos y rehacer el análisis (véase sección "Transformaciones para estabilizar varianza" del capítulo 5).
   e) Determinar cuál es la respuesta esperada en el mejor tratamiento.

10. *Hacer corridas confirmatorias del proceso en el mejor tratamiento*. Emplear suficientes repeticiones para que se tenga una buena estimación de lo logrado.

Conclusiones finales

11. Cerrar y concluir el proyecto adecuadamente. Para ello se recomienda realizar las siguientes actividades:

    a) Decidir qué medidas implementar para generalizar el resultado del estudio y garantizar que las mejoras se mantengan. Quizá sea conveniente aplicar las mismas medidas en un proceso similar, modificar procesos, procedimientos e instrucciones de trabajo y establecer medidas de control (véase Gutiérrez Pulido y De la Vara, 2004).

b) Hacer un resumen de los principales resultados logrados desde el punto de vista técnico.
c) Hacer una evaluación de los logros obtenidos con el proyecto.
d) Diseñar una presentación para la difusión del proyecto. Esto ayuda a reforzar el trabajo y a reconocer los logros.

Control de factores de bloque y de ruido

Hemos hecho énfasis en que cuando se planea un experimento es preciso contemplar todos los aspectos que podrían influir en el resultado, así como las fuentes de variación o factores que, aun cuando no sean de interés estudiar, se piensa que sí pueden influir en la variable de respuesta. Entonces hay tres maneras de controlar el efecto de este tipo de factores, que consisten en:

1. *Estudiarlo o bloquearlo activamente durante el experimento*. En este caso, por lo general se asignan varios niveles al factor de bloque y en cada nivel (o bloque) se prueban todos o algunos de los tratamientos en orden aleatorio. A la hora del análisis se puede obtener una medida del efecto del factor de bloque.

   El control directo de un factor de bloque como un factor más, es otra posibilidad. Sin embargo, no es económicamente factible ni vale la pena estudiar de esta forma todos los factores de bloque posibles. En la práctica sólo se estudian aquellos factores de bloque que pueden influir en la comparación y es económicamente factible fijarlos en diferentes niveles. Por ejemplo, se suelen estudiar en forma directa factores de bloque como: operadores, tipos de producto, lotes, turnos, tipos de material, etc. Los diseños factoriales que se analizaron en los capítulos anteriores pueden correrse en bloques (véase capítulo 6). El apoyo de un software es imprescindible.

2. *Nulificar su efecto al mantenerlo fijo en su valor usual durante todo el estudio*. Si el factor de bloque son los operadores de las máquinas que se quieren comparar, la estrategia equivaldría a utilizar el mismo operador. En este caso, se tendría la limitante de que los resultados son válidos sólo para ese operador, y no se eliminaría del todo el efecto del operador porque éste puede tener su máquina preferida.

   Otro ejemplo de esta estrategia es mantener fijo, durante el experimento, algún parámetro del proceso porque se considera que no vale la pena estudiar su efecto. El mantenerlo fijo implica verificar que en realidad se mantiene fijo durante el estudio; no basta decir que lo está porque la palanca de control no se mueve. Es decir, se requiere vigilancia especial (no la de diario) para comprobar que en realidad se mantiene fijo.

3. *Para nulificar en lo posible el efecto indeseable de factores de ruido, como es el caso de las variables ambientales, se aplica el principio de aleatorización*. Este principio hace que el posible efecto de este factor se reparta "equitativamente" en todos los tratamientos. Al nulificar un factor de ruido

de esta manera, no es posible medir su posible efecto, pero éste es el último y único recurso cuando es muy costoso mantenerlo fijo o considerar varios niveles en dicho factor.

Un ejemplo de factores de ruido que suelen controlarse mediante la aleatorización de las pruebas son las variables ambientales, como temperatura ambiental, humedad ambiental, presión, etc. Otros factores que también se controlan de este modo son: cansancio de los operadores, calentamiento del equipo y en general todas las variables que no podemos evitar que se muevan solas con el transcurso del tiempo.

Cabe recordar que no es objetivo del experimento estudiar el efecto de los factores de bloque, ya que estos factores sólo se controlan por la necesidad de lograr una comparación justa y precisa de los tratamientos. Cuando se decide no controlar de ningún modo un factor de bloque y se deja actuar libremente durante el experimento por considerar que no tiene efecto, se convierte en un factor de ruido que meterá algún efecto al experimento, que esperamos sea pequeño.

Qué sigue después del primer experimento

Muchas veces, aunque el experimento se haya planeado y realizado bien, éste no es definitivo ni concluyente en el sentido de que siempre quedan cuestiones pendientes, surgen nuevas preguntas o hipótesis acerca del problema. Incluso, pueden surgir dudas sobre lo adecuado de las acciones efectuadas. El camino a seguir después del primer experimento se basa en los resultados obtenidos en este primer estudio. Existen al menos tres tipos de acciones a seguir que se describen a continuación:

1. *El estudio se planea para desarrollarse por etapas, ya sea por la gran cantidad de factores que se tienen al principio o porque la metodología así lo propone*. Puede ocurrir que al comienzo de la investigación se tienen demasiados factores, y entonces es difícil pensar que el primer experimento pueda ser definitivo. La estrategia inicial en este caso es correr un diseño altamente fraccionado o *saturado* (que permita estudiar sólo los efectos principales e ignore las interacciones) para detectar los pocos factores que tienen mayor influencia, y posteriormente plantear un segundo experimento más completo con éstos.
2. *El estudio se planea para desarrollarse por etapas, aunque no sean muchos los factores que se tienen desde el principio*. Si se tiene una cantidad moderada de factores, digamos de cinco a ocho, y como cada corrida es costosa, se quiere evitar un exceso de pruebas. La estrategia recomendada es correr inicialmente un diseño factorial fraccionado de resolución III o IV, y después, si hubiera necesidad de aclarar la confusión de los efectos, se corre otra fracción adicional para eliminar dichas ambigüedades. Otro ejemplo de esta situación es cuando se tienen pocos factores (pensemos en tres) y se decide correr inicialmente sólo una réplica del experimento, pero a la hora del análisis se encuentra que algunos de los efectos quedan intermedios y no es fácil decidir si considerados significativos o no. Para definir

esta situación se decide correr una réplica adicional para dar más potencia al experimento y tener más grados de libertad para el error. Así, las ocho corridas adicionales sólo se llevan a cabo cuando son necesarias.

3. *Se comete algún error de planeación*. En ocasiones, la necesidad de volver a experimentar se debe a una planeación inadecuada del primer experimento. Un error frecuente es dedicar poco tiempo a la etapa de planeación.

Por ejemplo, un error crítico y muy común es seleccionar de manera inadecuada el *ancho entre los niveles* de los factores, que sólo se corrige corriendo otra vez el experimento o una parte de éste. Otro error es *no haber incluido o no controlar adecuadamente un factor* que tiene un efecto potencialmente importante. Es necesario aprender de este tipo de errores y diseñar un nuevo experimento de una mejor manera.

En la figura 10.2 se representan las alternativas típicas a seguir después de un primer experimento. Para fines de la representación de las diferentes acciones, se utilizan tres factores, pero debe entenderse que varias de estas acciones tienen mayor utilidad con una cantidad moderada de factores ($5 < k \leq 8$). A continuación se describen cada una de estas posibles acciones:

a) *Agregar otra fracción en un diseño factorial* 2^{k-p}. Esto se hace para eliminar algunas dudas o ambigüedades que surgen al interpretar efectos alias en diseños de resolución III y IV. Recordemos del capítulo 8, que dos efectos alias son inseparables, al menos que se obtenga una fracción adicional que, al combinarse con la primera, permita separar los efectos.

b) *Reescalar*. Esta acción se aplica cuando no se tomó el espaciamiento adecuado entre los niveles en uno o varios de los factores controlados en el experimento y, por lo tanto, no se pudo detectar el efecto de esos factores. Corregir esta falla en la planeación implica volver a correr al menos la mitad del experimento, y a veces por completo, cuando no se logran reproducir las condiciones de operación iniciales.

c) *Quitar y agregar factores*. En un segundo experimento se quita un factor cuando se ha comprobado que no afecta de ningún modo a la respuesta, y se puede agregar otro que no se había considerado por alguna razón pero que ahora interesa estudiar su efecto. Es recomendable que, desde el primer experimento, no se deje ningún factor controlable fuera del estudio si éste tiene posibilidades de afectar ya sea solo o interactuando. La acción de *colapsar* es una manera elegante de eliminar del análisis de un experimento un factor que no afecta (ver capítulo 6).

d) *Repetir*. Se recomienda replicar o repetir el experimento, al menos parcialmente, cuando hubo efectos para los cuales no se pudo ser concluyente, en el sentido de que no quedó claro si afectan o no. En el ANOVA, estos efectos se distinguen porque sus valores-*p* son mayores que 0.05 pero a la vez pequeños, por ejemplo, son menores que 0.1. El Pareto de efectos las barras correspondientes a tales efectos tienen longitud intermedia, y entonces surge la pregunta: ¿falta potencia a la prueba? Una manera de aclarar esta duda es correr más repeticiones o réplicas (una más) del experimento. Las repeticiones auténticas permiten tener un mejor estimador del error y mejorar sensiblemente el análisis de varianza.

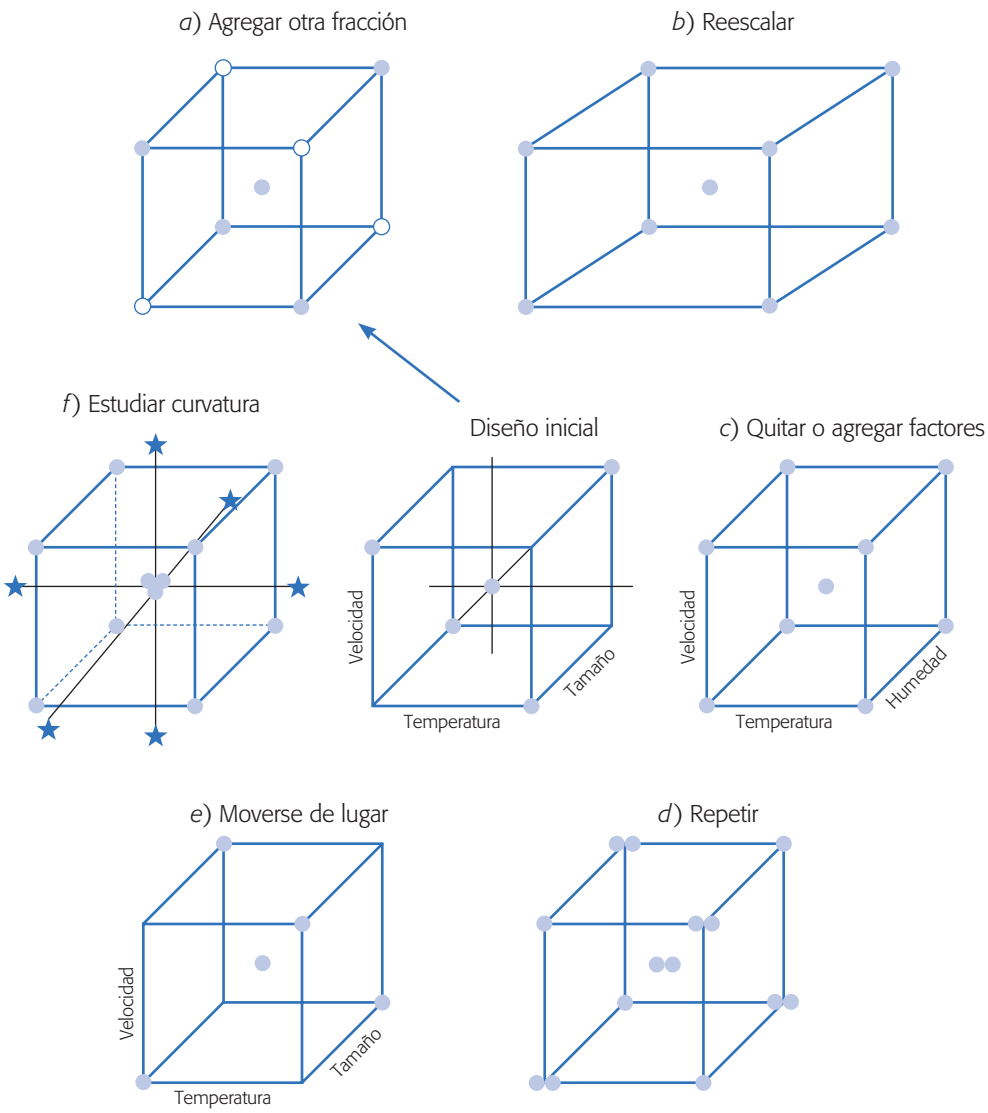


Figura 10.2 Posibles acciones después del primer experimento.

e) *Moverse de lugar*. Esta acción y la siguiente conforman la estrategia experimental llamada *metodología de superficie de respuesta*, la cual se presenta en el capítulo 12. Ambas tienen que ver con el interés, por parte del experimentador, de llegar a las condiciones óptimas de operación del proceso. *Donde* no basta con obtener un *mejor tratamiento* o *tratamiento elegido* en el primer experimento, sino que se buscan mejores maneras de operar el proceso. Del primer experimento, mediante el modelo de regresión ajustado, se pueden sacar conclusiones acerca de hacia qué dirección mover los niveles de cada factor en aras de un mejor desempeño de la variable de interés. Esta estrategia generalmente se trabaja con pocos factores (no más de cuatro), corriendo en el primer experimento los puntos necesarios para ajustar un *modelo de primer orden*.

Modelo de primer orden
Modelo estadístico que contiene sólo los términos correspondientes a los efectos principales de los factores. Su representación geométrica es un plano o hiperplano.

f) *Aumentar*. Cuando se busca optimizar un proceso, y en el primer experimento se detectó curvatura pura y/o la superficie de respuesta no es descrita adecuadamente por un modelo de primer orden, entonces será necesario *aumentar el experimento*. Para ello, se seleccionan en forma adecuada puntos adicionales y se corren, a fin de estudiar no sólo efectos principales e interacciones dobles, sino también los efectos cuadráticos puros. La característica fundamental del diseño aumentado es que tiene al menos tres niveles en cada factor para estudiar ese tipo de efectos. El diseño aumentado que se dibuja en la figura 10.2 se llama *diseño de composición central* o *diseño central compuesto* (capítulo 12).

Qué hacer cuando ningún efecto es significativo

Cuando en un estudio experimental ningún o casi ninguno de los factores estudiados fue significativo, se debe sacar provecho de lo realizado y tratar de entender por qué ningún efecto fue significativo, a pesar de que las conjeturas señalaban que sí. Esto se refleja en un R^2 bajo o muy bajo, o en que en el ANOVA casi ningún efecto resultó significativo. Algunas de las razones, no excluyentes unas de las otras, por las que podría pasar lo anterior son:

1. El proceso opera con una alta variación, de tal forma que el cuadrado medio del error es grande y, en consecuencia, prácticamente ningún factor resultó activo (recordemos que para que un efecto sea activo necesita que su cuadrado medio sea por los menos unas tres o cuatro veces más grande que el cuadrado medio del error). En ese caso se debe tratar de revalorar la situación y ver qué fuentes de variación que se fueron al error pudieron actuar durante el experimento. En caso de encontrar algunas, es preciso ver la posibilidad de consideradas en un segundo experimento, ya sea manteniéndolas fijas, bloqueándolas, como factor o con algún tipo de apareamiento (véase ejemplo de la sección "Poblaciones pareadas" del capítulo 2).
2. Los niveles asignados a los factores fueron muy estrechos, de forma que la diferencia entre lo que pasa en un nivel y otro es prácticamente imperceptible por la variable de respuesta. En este caso habrá que reescalar y volver a correr el experimento. En teoría, cuando los factores sí influyen y se eligen bien sus niveles, la dispersión de la variable de respuesta durante el experimento debe ser mayor que la variación durante la operación normal del proceso. Una forma de verificar si los niveles fueron lo suficientemente amplios, consiste en comparar la variación observada en los datos o experimentos contra la variación normal que se tiene en el proceso, y si la primera no es evidentemente más grande, entonces eso puede ser un indicio de que los niveles no fueron elegidos en forma adecuada.
3. Los factores estudiados no son los principales responsables de la variación de la variable de respuesta; por lo tanto, hay otras situaciones o factores no considerados que son los que en realidad tienen influencia sobre la respuesta.

En este caso, eso puede ser un aprendizaje importante, ya que ayudará a enfocar la solución desde otra perspectiva.

4. Se observó poca variación en la variable de respuesta. Esto ocurre cuando las pruebas realizadas no lograron "mover" más de lo usual a la variable de respuesta, y eso puede ocurrir por cualquiera de las dos situaciones anteriores, o porque el tamaño de la prueba fue insuficiente para detectar los cambios de interés. Esto último ocurre sobre todo cuando la variable de respuesta implica algún tipo de conteo, por ejemplo, el porcentaje de artículos defectuosos en cada corrida, y ese porcentaje normalmente es pequeño, así que será necesario producir en cada prueba muchas piezas para que se pueda observar algún artículo defectuoso. En la ecuación 6.13 del capítulo 6 se da una fórmula para determinar el tamaño de la prueba en función de la magnitud del problema, y en los ejercicios 9, 14 y 17 del mismo capítulo se muestran casos en los que se determinó el tamaño de prueba. Por lo tanto, si el tamaño de prueba no fue suficiente, entonces habrá que volver a correr el experimento para agregar mayor información a la ya disponible.

A manera de resumen de la presente sección, el experimentador siempre debe tener presente el hecho de que todo experimento genera información y aprendizaje, lo cual será útil para reorientar la búsqueda, para no cometer los mismos errores en un siguiente estudio experimental o para afinar las conjeturas que se tienen sobre el problema. En este sentido, se debe tener mucha precaución de que si no se obtuvo la respuesta buscada, no se confirmó la conjetura que se tenía o si el R^2 en el análisis de varianza fue pequeño, entonces no se debe cometer el error de desechar el experimento y valorarlo como un fracaso. Por el contrario, es preciso analizar qué pasó, así como reflexionar e investigar cuáles de las cuatro posibilidades anteriores pudieron ocurrir en el experimento. En otras palabras, en un estudio experimental siempre se debe tener presente el ciclo de Deming, además de sacar enseñanzas y conclusiones para usarlas en un nuevo ciclo.

Preguntas y ejercicios

1. Explique cuáles son las principales actividades a realizar en la etapa de planeación de un estudio experimental.
2. Apoyándose en las ideas del ejemplo 10.1 y el paso 5 de la sección "Etapas y actividades de la planeación" de este capítulo, describa lo que se debe hacer cuando en un estudio factorial hay complicaciones para aleatorizar completamente el orden de las corridas debido a las limitaciones que impone un factor.
3. Si en un estudio factorial existe un factor que se desea estudiar, pero es muy difícil mantenerlo fijo en un nivel determinado, más bien, está variando alrededor del nivel que se fija. ¿Qué se recomienda hacer en estos casos para estudiar tal factor?
4. Describa el ciclo de Deming en el contexto de un estudio experimental.
5. Una vez que en un experimento factorial se detectó el mejor tratamiento y se proponen las condiciones de operación futura del proceso, se recomienda hacer corridas y pruebas confirmatorias. ¿En qué consisten éstas y por qué es necesario realizarlas?

6. Si en un estudio factorial se concluye que sólo tres de los cinco factores estudiados impactan de alguna manera la variable de respuesta, y los otros dos no influyen en ésta, ¿cómo se deben elegir los niveles de los factores que no son significativos?
7. Ilustre para un factor en particular, qué significa que sus niveles estén demasiado estrechos y explique por qué esto representa un problema en un estudio experimental.
8. En todo estudio factorial participan los factores y la o las variables de respuesta. ¿En estos casos quién es medio y quién es el fin último? Explique su respuesta.
9. ¿Qué se recomienda hacer con un factor o fuente de variación, que aunque no interesa estudiar, se piensa que sí influye en la variable de respuesta?
10. Si en un estadio factorial al principio se tienen muchos factores, más de cinco, ¿qué tipo de diseño se recomienda aplicar?
11. En el caso de la pregunta anterior, ¿qué se recomienda hacer en un segundo experimento?
12. Cuando en un estudio factorial ninguno o casi ningún factor fue significativo y/o el R^2 es bajo, ¿será un indicativo de que el experimento fue un fracaso? Argumente ampliamente.
13. ¿Cuáles son las principales razones por las que en un estudio factorial ninguno o casi ningún factor fue significativo y/o el R^2 es bajo?

Capítulo 11
Análisis de regresión

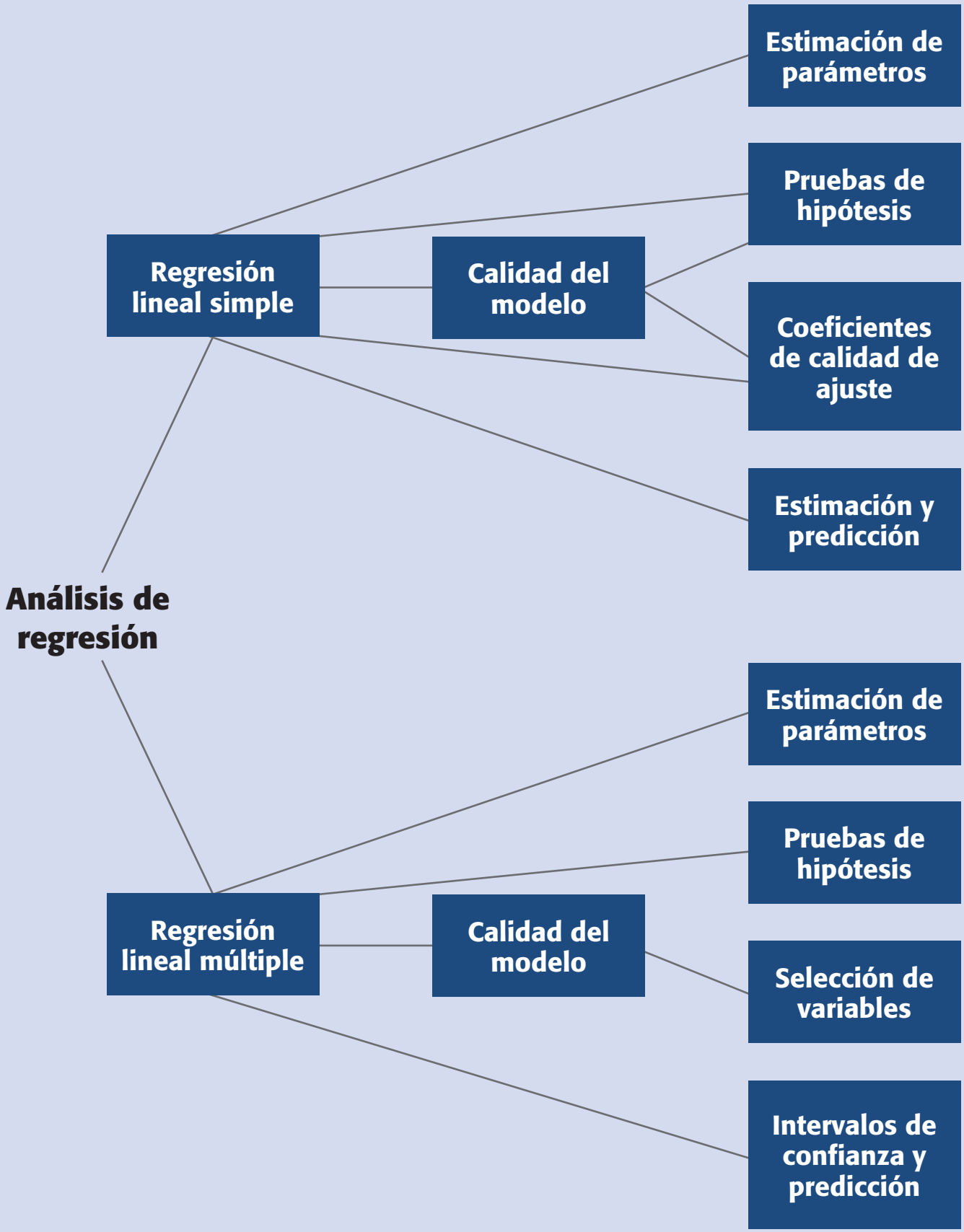
Sumario

- Regresión lineal simple
- Pruebas de hipótesis en la regresión lineal simple
- Calidad del ajuste en regresión lineal simple
- Estimación y predicción por intervalo en regresión simple
- Regresión lineal múltiple
- Pruebas de hipótesis en regresión lineal múltiple
- Intervalos de confianza y predicción en regresión múltiple
- Uso de un software estadístico

Objetivos de aprendizaje

- Entender el papel del análisis de regresión dentro de los diseños experimentales.
- Aplicar las pruebas de hipótesis en la regresión lineal y evaluar la calidad de un modelo.
- Diferenciar entre regresión lineal simple y múltiple, y aplicar cada una al caso apropiado.
- Utilizar de manera correcta un modelo de regresión para propósitos de estimación y predicción.

Mapa conceptual



Conceptos clave

- Análisis de regresión
- Coeficiente de correlación
- Coeficiente de correlación múltiple
- Coeficiente de determinación R^2
- Error estándar de estimación
- Media del error absoluto
- Método de mínimos cuadrados
- Prueba de Durbin-Watson
- Prueba de falta de ajuste
- Residuos
- Suma de cuadrados del error

Análisis de regresión
Explica en forma matemática el comportamiento de una variable de respuesta en función de una o más variables independientes.

El *análisis de regresión* tiene como objetivo modelar en forma matemática el comportamiento de una variable de respuesta en función de una o más variables independientes (factores). Por ejemplo, suponga que el rendimiento de un proceso químico está relacionado con la temperatura de operación. Si mediante un modelo matemático es posible describir tal relación, entonces este modelo puede ser usado para propósitos de predicción, optimización o control.

Para estimar los parámetros de un *modelo de regresión* son necesarios los datos, los cuales pueden obtenerse de experimentos planeados, de observaciones de fenómenos no controlados o de registros históricos.

Regresión lineal simple

Sean dos variables X y Y, suponga que se quiere explicar el comportamiento de Y con base en los valores que toma X. Para esto, se mide el valor de Y sobre un conjunto de n valores de X, con lo que se obtienen n parejas de puntos $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. A Y se le llama la *variable dependiente* o la *variable de respuesta* y a X se le conoce como *variable independiente* o *variable regresora*. La variable X no necesariamente es aleatoria, ya que en muchas ocasiones el investigador fija sus valores; en cambio, Y sí es una variable aleatoria. Una manera de estudiar el comportamiento de Y con respecto a X es mediante un modelo de regresión que consiste en ajustar un modelo matemático de la forma:

$$Y = f(X)$$

a las n parejas de puntos. Con ello, se puede ver si dado un valor de la variable independiente X es posible predecir el valor promedio de Y.

Suponga que las variables X y Y están relacionadas linealmente y que para cada valor de X, la variable dependiente, Y, es una variable aleatoria. Es decir, que cada observación de Y puede ser descrita por el modelo:

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad \textbf{(11.1)}$$

donde ε es un error aleatorio con media cero y varianza σ^2. También suponga que los errores aleatorios no están correlacionados. La ecuación (11.1) es conocida como el *modelo de regresión lineal simple*.[1] Bajo el supuesto de que este modelo es adecuado

[1] Existen otros modelos de regresión que sólo incluyen una variable independiente y que se aplican cuando se espera o se observa que la relación entre X y Y no es modelada por una línea recta. Algunos de estos modelos son los siguientes:

El exponencial: $Y = e^{\beta_0 + \beta_1 X}$

El recíproco-Y: $Y = 1/(\beta_0 + \beta_1 X)$

Mutiplicativo: $Y = \beta_0 X^{\beta_1}$

Recíproco-X: $Y = \beta_0 + \beta_1 / X$

Logarítmico-X: $Y = \beta_0 + \beta_1 \ln(X)$

Raíz cuadrada-Y: $\sqrt{Y} = \beta_0 + \beta_1 X$

y como el valor esperado del error es cero, $E(\varepsilon) = 0$, se puede ver que el valor esperado de la variable Y, para cada valor de X, está dado por línea recta

$$E(Y|X) = \beta_0 + \beta_1 X \tag{11.2}$$

en donde β_0 y β_1 son los parámetros del modelo y son constantes desconocidas. Por lo tanto, para tener bien especificada la ecuación que relaciona las dos variables será necesario estimar los dos parámetros, que tienen los siguientes significados: β_0 es el punto en el cual la línea recta intercepta o cruza el eje y, y β_1 es la pendiente de la línea, es decir, es la cantidad en que se incrementa o disminuye la variable Y por cada unidad que se incrementa X.

Ejemplo 11.1

En un laboratorio se quiere investigar la forma en que se relaciona la cantidad de fibra (madera) en la pulpa con la resistencia del producto (papel). Los datos obtenidos en un estudio experimental se muestran en la tabla 11.1.

Es claro que la variable de respuesta o variable dependiente es la resistencia, por eso se denota con Y. Para tener una idea de la relación que existe entre X y Y, los 14 pares de datos son graficados en el diagrama de dispersión de la figura 11.1. Se observa que entre X y Y existe una correlación lineal positiva, ya que conforme aumenta X también se incrementa Y, por lo que es razonable suponer que la relación entre X y Y la explique un modelo de regresión lineal simple. Así, cada observación de Y, la podemos expresar como:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{11.3}$$

con $i = 1, 2, \ldots, n$ ($n = 14$). Para estimar β_0 y β_1 ajustamos la recta que explique de mejor manera el comportamiento de los datos en el diagrama de dispersión de la figura 11.1. En otras palabras, debemos encontrar la recta que pasa más cerca de todos

Tabla 11.1 Datos de resistencia de la pulpa, ejemplo 11.1.

Porcentaje de fibra	Resistencia
4	134
6	145
8	142
10	149
12	144
14	160
16	156
18	157
20	168
22	166
24	167
26	171
28	174
30	183

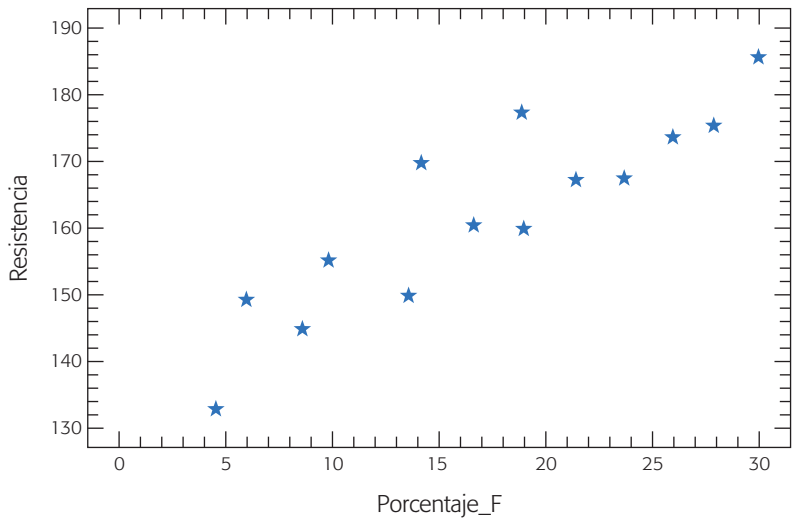


Figura 11.1 Diagrama de dispersión para los datos de resistencia de la pulpa.

Método de mínimos cuadrados
Procedimiento para estimar los parámetros de un modelo de regresión que minimiza los errores de ajuste del modelo.

los puntos. Un procedimiento para ajustar la mejor recta y, por lo tanto, para estimar β_0 y β_1 es mediante el *método de mínimos cuadrados*, el cual consiste en lo siguiente: si de la ecuación (11.3) despejamos los errores, los elevamos al cuadrado y los sumamos, obtenemos lo siguiente:

$$S = \sum_{i=1}^{n} (\varepsilon_i)^2 = \sum_{i=1}^{n} \left(y_i - [\beta_0 + \beta_1 x_i]\right)^2 \qquad \textbf{(11.4)}$$

De esta forma, se quieren encontrar los valores de β_0 y β_1 que minimizan la suma de los errores cuadrados. Es decir, se busca ajustar la recta de manera que la suma de las distancias en forma vertical de los puntos a la recta se minimice, como se ilustra en la figura 11.2.

El procedimiento matemático para minimizar los errores de la ecuación (11.4) y así encontrar los estimadores de mínimos cuadrados de β_0 y β_1, consiste en derivar a S con respecto a β_0, $\frac{\partial S}{\partial \beta_0}$ y derivar también a S respecto a β_1, $\frac{\partial S}{\partial \beta_1}$, se obtiene:

$$\frac{\partial S}{\partial \beta_0} = -\sum_{i=1}^{n} 2\left(y_i - [\beta_0 + \beta_1 x_i]\right)$$

$$\frac{\partial S}{\partial \beta_1} = -\sum_{i=1}^{n} 2x_i\left(y_i - [\beta_0 + \beta_1 x_i]\right)$$

Al igualar a cero las dos ecuaciones y resolverlas en forma simultánea con respecto a las dos incógnitas (β_0 y β_1), se obtiene la solución única:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \qquad \textbf{(11.5)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \textbf{(11.6)}$$

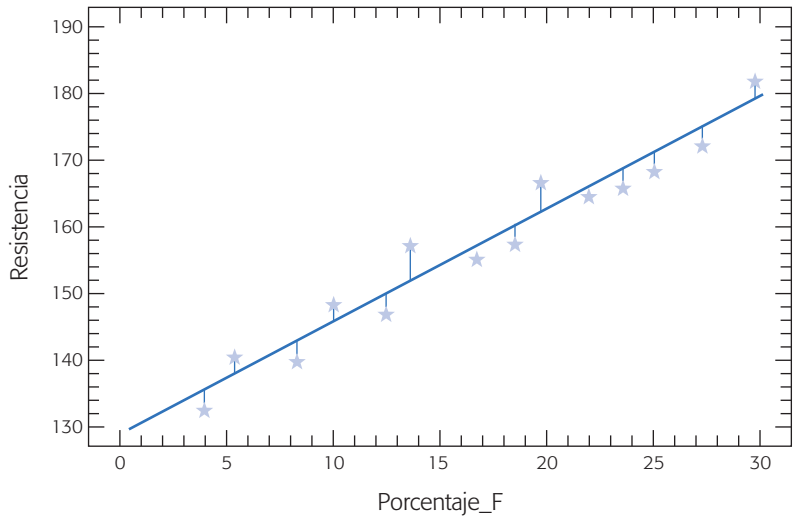


Figura 11.2 Minimizar la distancia vertical de los puntos a la recta.

donde

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{(\Sigma_{i=1}^{n} x_i)(\Sigma_{i=1}^{n} y_i)}{n} \qquad \textbf{(11.7)}$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{(\Sigma_{i=1}^{n} x_i)^2}{n} \qquad \textbf{(11.8)}$$

$\bar{x}, \bar{y}$ son las medias muestrales de las dos variables, es decir,

$$\bar{x} = \frac{\Sigma_{i=1}^{n} x_i}{n}$$

$$\bar{y} = \frac{\Sigma_{i=1}^{n} y_i}{n}$$

De esta forma, para obtener la recta ajustada es necesario aplicar las fórmulas anteriores, lo cual es muy sencillo, como se muestra en la tabla 11.2 para los datos de la resistencia de la pulpa.

$$S_{xy} = 39\,150 - \frac{(238)(2\,216)}{14} = 1\,478.0$$

$$S_{xx} = 4\,956 - \frac{(238)^2}{14} = 910.0$$

$$\hat{\beta}_1 = \frac{1\,478.0}{910.0} = 1.6242$$

$$\hat{\beta}_0 = 158.286 - (1.6242)(17) = 130.67$$

Por lo tanto, la línea recta que mejor explica la relación entre porcentaje de fibra y resistencia del papel, está dada por

$$\hat{Y} = 130.67 + 1.6242X \qquad \textbf{(11.9)}$$

Tabla 11.2 Procedimiento para realizar los cálculos para la regresión simple.

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
4	134	16	17 956	536
6	145	36	21 025	870
8	142	64	20 164	1 136
10	149	100	22 201	1 490
12	144	144	20 736	1 728
14	160	196	25 600	2 240
16	156	256	24 336	2 496
18	157	324	24 649	2 826
20	168	400	28 224	3 360
22	166	484	27 556	3 652
24	167	576	27 889	4 008
26	171	676	27 241	4 446
28	174	784	30 276	4 872
30	183	900	33 489	5 490
$\sum x_i = 238$	$\sum y_i = 2\,216$	$\sum x_i^2 = 4\,956$	$\sum y_i^2 = 353\,342$	$\sum x_i y_i = 39\,150$
$\bar{x} = 17$	$\bar{y} = 158.286$			

En la figura 11.2 se muestra el ajuste de esta línea. De esta manera, por cada punto porcentual de incremento en el porcentaje de fibra, se espera un incremento de la resistencia de 1.6242 en promedio. La ecuación (11.9) sirve para estimar la resistencia promedio esperada para cualquier porcentaje de fibra utilizada, claro que esa estimación será más precisa en la medida que X esté dentro del intervalo de los valores con los que se hizo la estimación. Por ejemplo, para cada x_i, con el que se experimentó, se puede estimar el $\hat{y}_i$ con base en el modelo; la diferencia entre lo observado y lo estimado o predicho es una estimación del error ε_i. Tal estimación recibe el nombre de *residuo*, e_i, donde:

Residuos
Es la diferencia entre lo observado y lo estimado o predicho. Sirven para analizar el error de ajuste de un modelo.

$$e_i = y_i - \hat{y}_i$$

en la tabla 11.3 se muestran los residuos y predichos para el ejemplo 11.1. Más adelante veremos que estos residuos son de gran utilidad para verificar la calidad del ajuste del modelo.

Un aspecto que es importante resaltar es que los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ son variables aleatorias, ya que dependen de los valores observados, y_i. De esta manera, es posible evaluar el valor esperado y la varianza de los estimadores. Al respecto, es fácil demostrar los siguientes resultados:

$$E(\hat{\beta}_1) = E\left(\frac{S_{xy}}{S_{xx}}\right) = \beta_1 \qquad \textbf{(11.10)}$$

$$E(\hat{\beta}_0) = E(y - \hat{\beta}_1 \bar{x}) = \beta_0 \qquad \textbf{(11.11)}$$

$$V(\hat{\beta}_1) = V\left(\frac{S_{xy}}{S_{xx}}\right) = \frac{\sigma^2}{S_{xx}} \qquad \textbf{(11.12)}$$

Tabla 11.3 Valores ajustados, $\hat{y}_i$, y residuos, e_i, para el ejemplo 11.1.

x_i	y_i	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$
4	134	137.2	–3.2
6	145	140.4	4.6
8	142	143.7	–1.7
10	149	146.9	2.1
12	144	150.2	–6.2
14	160	153.4	6.6
16	156	156.7	–0.7
18	157	159.9	–2.9
20	168	163.2	4.8
22	166	166.4	–0.4
24	167	169.7	–2.7
26	171	172.9	–1.9
28	174	176.2	–2.2
30	183	179.4	3.6

$$V(\hat{\beta}_0) = V(\bar{y} - \hat{\beta}_1\bar{x}) = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right] \tag{11.13}$$

De esta manera $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores insesgados de los correspondientes parámetros. Además, $\hat{\beta}_0$ y $\hat{\beta}_1$ como variables aleatorias no son independientes, y su covarianza está dada por

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2\bar{x}}{S_{xx}} \tag{11.14}$$

Dadas las expresiones para las varianzas de $\hat{\beta}_0$ y $\hat{\beta}_1$, se aprecia que para estimarlas es necesario calcular σ^2 que, como señalamos en la expresión (11.1), σ^2 es la varianza del error aleatorio, ε. Así que es natural utilizar los residuos para hacer tal estimación. Para ello, la suma de cuadrados de los residuos o *suma de cuadrados del error* está dada por:

Suma de cuadrados del error
Es la suma de los residuos al cuadrado, y se utiliza para estimar la varianza del error de ajuste de un modelo.

$$SC_E = \sum_{i=1}^{n} e_1^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n} \left(y_i - [\hat{\beta}_0 + \hat{\beta}_i x_i]\right)^2 \tag{11.15}$$

$$= S_{yy} - \hat{\beta}_1 S_{xy} = S_{yy} - \frac{(S_{xy})^2}{S_{xx}}$$

donde:

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{(\Sigma_{i=1}^{n}\, y_i)^2}{n} \tag{11.16}$$

A partir de la ecuación (11.15) se obtiene que el valor esperado de la suma de cuadrado del error está dado por:

$$E(SC_E) = (n-2)\sigma^2 \tag{11.17}$$

Por lo tanto, un estimador insesgado de σ^2 está dado por:

$$\hat{\sigma}^2 = \frac{SC_E}{n-2} = CM_E \tag{11.18}$$

Como se aprecia en la expresión anterior, $\hat{\sigma}^2$ recibe el nombre de *cuadrado medio del error*; y la raíz cuadrada de éste, es decir, $\hat{\sigma} = \sqrt{CM_E}$, se conoce como *error estándar de estimación.*

Pruebas de hipótesis en la regresión lineal simple

En cualquier análisis de regresión no basta hacer los cálculos que se explicaron antes, sino que es necesario evaluar qué tan bien el modelo (la línea recta) explica la relación entre X y Y. Una primera forma de hacer esto es probar una serie de hipótesis (véase capítulo 2) sobre el modelo. Para ello es necesario suponer una distribución de probabilidad para el término de error, ε_i. Es usual suponer normalidad: ε_i se distribuye en forma normal, independiente, con media cero y varianza σ^2 ($\varepsilon_i \sim NID(0, \sigma^2)$).

Por lo general, la hipótesis de mayor interés plantea que la pendiente es significativamente diferente de cero. Esto se logra al probar la siguiente hipótesis:

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_A &: \beta_1 \neq 0 \end{aligned} \tag{11.19}$$

Para encontrar el estadístico de prueba para esta hipótesis, es fácil ver que $\hat{\beta}_1$ sigue una distribución normal, y dado que la media y varianza de $\hat{\beta}_1$ están dadas por las ecuaciones (11.10) y (11.12), respectivamente. Entonces una estimación de $V(\hat{\beta}_1)$ está dada por:

$$\hat{V}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}} = \frac{CM_E}{S_{xx}} \tag{11.20}$$

La cantidad $\sqrt{\hat{V}(\hat{\beta}_1)}$ es la desviación estándar del estimador y recibe el nombre de *error estándar* de $\hat{\beta}_1$. Si la hipótesis nula es verdadera el siguiente estadístico

$$t_0 = \frac{\hat{\beta}_1}{\sqrt{CM_E / S_{xx}}} \tag{11.21}$$

tiene una distribución *t-Student* con $n - 2$ grados de libertad. En el capítulo 2 vimos tres maneras diferentes de rechazar una hipótesis con base en un estadístico de prueba. Al aplicar uno de estos criterios, se rechaza H_0 si el valor absoluto de este estadístico es mayor que el correspondiente valor crítico obtenido de tablas, es decir, se rechaza H_0 si:

$$|t_0| > t_{(\alpha/2, n-2)} \tag{11.22}$$

En caso contrario no se rechaza H_0. No rechazar que $\beta_1 = 0$, en el caso del modelo de regresión lineal simple, implica que no existe una relación lineal significativa entre X y Y; por lo tanto, no existe relación entre estas variables o ésta es de otro tipo. En ocasiones, en lugar de probar que $\beta_1 = 0$, puede ser de interés probar que es igual a cierta constante ($H_0 : \beta_1 = c$), en ese caso en el numerador del estadístico de la expresión (11.21) se resta c, es decir, el estadístico queda de la siguiente manera $t_0 = (\hat{\beta}_1 - c)/\sqrt{CM_E/S_{xx}}$, y el criterio de rechazo es el mismo.

Si se utiliza como criterio de rechazo la comparación de la significancia observada (*p-value* o valor-*p*) contra la significancia predefinida (α), entonces se *rechaza* H_0 si valor-$p < \alpha$.

Por otro lado, con respecto del parámetro β_0 suele ser de interés probar la siguiente hipótesis:

$$\begin{aligned} H_0 &: \beta_0 = 0 \\ H_A &: \beta_0 \neq 0 \end{aligned} \tag{11.23}$$

Para encontrar el estadístico de prueba para esta hipótesis, es fácil ver que $\hat{\beta}_0$ sigue una distribución normal, y dado que la media y varianza de $\hat{\beta}_0$ están dadas por las ecuaciones (11.11) y (11.13), respectivamente; entonces, una estimación de $V(\hat{\beta}_0)$ está dada por:

$$\begin{aligned} \hat{V}(\hat{\beta}_0) &= \hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right] \\ &= CM_E \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right] \end{aligned} \tag{11.24}$$

Si la hipótesis nula es verdadera, el siguiente estadístico:

$$t_0 = \frac{\hat{\beta}_0}{\sqrt{CM_E \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}} \tag{11.25}$$

tiene una distribución *t-Student* con $n - 2$ grados de libertad, por lo que H_0 se rechaza si:

$$|t_0| > t_{(\alpha/2, n-2)} \tag{11.26}$$

o si se utiliza el criterio de la significancia observada se *rechaza* H_0 si valor-$p < \alpha$. No rechazar que $\beta_0 = 0$ simplemente significa que el punto de corte de la línea recta pasa por el origen, es decir, pasa por (0, 0). En ocasiones, en lugar de probar que $\beta_0 = 0$, puede ser de interés probar que es igual a cierta constante ($H_0 : \beta_0 = c$); en ese caso, en el numerador del estadístico de la expresión (11.25) se resta c, es decir, el estadístico queda de la siguiente manera: $t_0 = (\hat{\beta}_0 - c)/\sqrt{CM_E \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}$ y el criterio de rechazo es el mismo.

La estimación de los parámetros del modelo y las pruebas de hipótesis sobre los mismos se sintetizan en la tabla 11.4.

Tabla 11.4 Análisis de regresión para el modelo $Y = \beta_0 + \beta_1 X$.

Parámetro	Estimación	Error estándar	Estadístico	Valor-*p*
Intercepción	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$	$\sqrt{CM_E\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}$	$\frac{\hat{\beta}_0}{\sqrt{CM_E\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}}$	$\Pr(T > \lvert t_0 \rvert)$
Pendiente	$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$	$\sqrt{CM_E / S_{xx}}$	$\frac{\hat{\beta}_1}{\sqrt{CM_E / S_{xx}}}$	$\Pr(T > \lvert t_0 \rvert)$

Estimación por intervalo para los parámetros de la regresión lineal simple

Con las pruebas de hipótesis que hemos descrito antes para los parámetros de la regresión lineal simple que se sintetizan en la tabla 11.4, y dada la relación entre prueba de hipótesis y estimación por intervalo (véase capítulo 2); se genera de manera directa la estimación por intervalo para los parámetros referidos. De esta forma, el intervalo de confianza al $100(1-\alpha)\%$, para β_1 está dado por:

$$\hat{\beta}_1 \pm t_{(\alpha/2, n-2)} \sqrt{CM_E / S_{xx}} \tag{11.27}$$

Mientras que el intervalo de confianza al $100(1-\alpha)\%$, para β_0 está dado por:

$$\hat{\beta}_0 \pm t_{(\alpha/2, n-2)} \sqrt{CM_E \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} \tag{11.28}$$

Análisis de varianza del modelo de regresión

Otro enfoque para analizar la significancia del modelo es descomponer la variabilidad observada, y a partir de ello probar hipótesis. Efectivamente, la variabilidad total observada en la variable de respuesta puede ser medida a través de S_{yy}, que está dada por la ecuación (11.16), pero también se puede descomponer de la siguiente manera:

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i) \tag{11.29}$$

El primer componente de S_{yy} se denota por $SC_R = \Sigma_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ y mide la variabilidad explicada por la recta de regresión (modelo de regresión), y se le conoce como la *suma de cuadrado de regresión*. Mientras que, como vimos en la ecuación (11.15), el segundo componente de S_{yy}, corresponde a la suma de cuadrados del error, $SC_E = \Sigma_{i=1}^{n} (y_i - \hat{y}_i)$, y mide la variabilidad no explicada por la recta de regresión (modelo de regresión). De esta manera, la ecuación (11.29) toma la siguiente forma:

$$S_{yy} = SC_R + SC_E \tag{11.30}$$

Los grados de libertad para S_{yy} son $n-1$, SC_R tiene un grado de libertad y SC_E tiene $n-2$. Al dividir las sumas de cuadrados entre sus grados de libertad obtenemos

los cuadrados medios: $CM_E = SC_E/(n-2)$ y $CM_R = SC_R/1$. Además de la ecuación (11.30) y de la expresión (11.15) se obtiene que la fórmula para calcular SC_R está dada por:

$$SC_R = \hat{\beta}_1 S_{xy} \tag{11.31}$$

Todo lo anterior podemos utilizarlo para generar otra forma de probar la hipótesis sobre la significancia de la regresión:

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_A &: \beta_1 \neq 0 \end{aligned} \tag{11.32}$$

ya que si H_0 es verdadera, entonces el siguiente estadístico:

$$F_0 = \frac{SC_R/1}{SC_E/(n-2)} = \frac{CM_R}{CM_E} \tag{11.33}$$

tiene una distribución F con 1 y $n-2$ grados de libertad en el numerador y denominador, respectivamente. Por lo tanto, se rechaza $H_0 : \beta_1 = 0$, si el estadístico de prueba es mayor que el valor crítico correspondiente, es decir, se rechaza H_0 si $F_0 > F_{(\alpha,\, 1,\, n-2)}$.

Esta forma de probar la significancia de la regresión, en el caso de la regresión lineal simple, es equivalente a la que se estableció a través del estadístico de la expresión (11.21), ya que al elevar éste al cuadrado obtenemos:

$$t_0^2 = \frac{\hat{\beta}_1^2 S_{xx}}{CM_E} = \frac{\hat{\beta}_1 S_{xy}}{CM_E} = \frac{CM_R}{CM_E} = F_0 \tag{11.34}$$

Esto se debe a que, en general, el cuadrado de una variable aleatoria *t-Student* con k grados de libertad, tiene una distribución $F_{(1,\, k)}$.

El análisis de varianza para probar la significancia del modelo de regresión se resume en la tabla 11.5.

Ejemplo 11.2

Al hacer el análisis de regresión y el análisis de varianza para el modelo que se ajustó a los datos del ejemplo 11.1 (resistencia de la pulpa), se obtienen las tablas 11.6 y 11.7. Veamos los detalles.

En la primera sección de este capítulo ya habíamos visto que $\hat{\beta}_0 = 130.675$ y $\hat{\beta}_1 = 1.62418$. Además de la tabla 11.2 se puede obtener:

Tabla 11.5 Análisis de varianza para el modelo de regresión simple.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor-p
Regresión	$SC_R = \hat{\beta}_1 S_{xy}$	1	CM_R	CM_R/CM_E	$\Pr(F > F_0)$
Error o residual	$SC_E = S_{yy} - \hat{\beta}_1 S_{xy}$	$n-2$	CM_E		
Total	S_{yy}	$n-1$			

Tabla 11.6 Análisis de regresión para el ejemplo 11.2.

Parámetro	Estimación	Error estándar	Estadístico	Valor-*p*
Intercepción	130.675	2.41779	54.0472	0.0000
Pendiente	1.62418	0.128504	12.6391	0.0000

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{(\Sigma_{i=1}^{n}\, y_i)^2}{n} = 353\,342 - \frac{(2\,216)^2}{14} = 2\,580.9$$

$$SC_E = S_{yy} - \hat{\beta}_1 S_{xy} = 2\,580.9 - (1.62418)(1\,478) = 180.36$$

de aquí,

$$CM_E = \frac{SC_E}{n-2} = \frac{180.36}{12} = 15.03$$

Los errores estándar:

$$\sqrt{\hat{V}(\hat{\beta}_0)} = \sqrt{CM_E\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} = \sqrt{15.03\left[\frac{1}{14} + \frac{(17)^2}{910.0}\right]} = 2.418$$

$$\sqrt{\hat{V}(\hat{\beta}_1)} = \sqrt{CM_E/S_{xx}} = \sqrt{(15.03)/(910.0)} = 0.12852$$

Así, los estadísticos para β_0 y para β_1 están dados, respectivamente, por:

$$\frac{\hat{\beta}_0}{\sqrt{CM_E\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}} = \frac{130.675}{2.418} = 54.043$$

$$\frac{\hat{\beta}_1}{\sqrt{CM_E/S_{xx}}} = \frac{1.62418}{0.12852} = 12.638$$

Con esto completamos la tabla 11.6, donde se muestra el análisis de regresión para la recta de regresión que relaciona el porcentaje de fibra con la resistencia de la pulpa. Como los valores *p* son menores que 0.05, entonces se rechazan las hipótesis nulas para ambos parámetros, por lo que se concluye que tanto β_0 como β_1 son significativamente diferentes de cero. Esto también se podría ver a través del método del valor crítico: de acuerdo con las tablas del apéndice el valor crítico de la distribución *t* de Student es: $t_{(0.025,12)} = 2.179$; además, es claro que ambos estadísticos de la tabla 11.6 son mayores que este valor crítico. De aquí que se llegue a la misma decisión que por el método del valor *p*.

Tabla 11.7 Análisis de varianza para el ejemplo 11.2.

Porcentaje de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor-*p*
Regresión	2 400.5	1	2 400.5	159.71	0.0000
Error o residuo	180.32	12	15.0271		
Total	2 580.86	13			

Prácticamente ya tenemos todos los cálculos para completar el análisis de varianza de la tabla 11.7, sólo restaría:

$$SC_R = \hat{\beta}_1 S_{xy} = (1.62418)(1\,478) = 2\,400.5$$

$$F_0 = \frac{CM_R}{CM_E} = \frac{2\,400.5}{15.03} = 159.71$$

En la tabla 11.7 se observa que el modelo de regresión es significativo, ya que el valor p es menor que 0.05. Esta misma conclusión se obtiene si se procede con el método del valor crítico, ya que de las tablas del apéndice se obtiene que $F_0 > F_{(0.05,\,1,\,12)} = 4.75$.

Calidad del ajuste en regresión lineal simple

En la sección anterior estudiamos pruebas de hipótesis para verificar que hay una relación significativa entre X y Y; sin embargo, no hemos visto si tal relación permite hacer estimaciones con una precisión aceptable. Por ejemplo, es de interés saber qué tanta de la variabilidad presente en Y fue explicada por el modelo, además si se cumplen los supuestos de los residuos.

Coeficiente de determinación R^2

Un primer criterio para evaluar la calidad del ajuste es observar la forma en que el modelo se ajustó a los datos. En el caso de la regresión lineal simple esto se distingue al observar si los puntos tienden a ajustarse razonablemente bien a la línea recta (véase figura 11.2). Pero otro criterio más cuantitativo es el que proporciona el *coeficiente de determinación*, que en el contexto de diseño de experimentos explicamos en el capítulo 6, y que ahora en el contexto de regresión está definido por:

Coeficiente de determinación R^2
Mide la proporción de la variabilidad en los datos (Y) que es explicada por el modelo de regresión.

$$R^2 = \frac{\text{Variabilidad explicada por el modelo}}{\text{Variabilidad total}} = \frac{SC_R}{S_{yy}} \qquad \textbf{(11.35)}$$

Es claro que $0 < R^2 \leq 1$. En general R^2 se interpreta como la proporción de la variabilidad en los datos (Y) que es explicada por el modelo. En el caso de los datos del ejemplo 11.1, a partir de la tabla 11.7 tenemos: $R^2 = (2\,400.5)/(2\,580.86) = 0.930$. Por lo tanto, podemos decir que 93% de la variación observada en la resistencia es explicada por el modelo (línea recta), lo cual nos dice que la calidad del ajuste es satisfactorio, y que por ello, la relación entre X y Y es descrita adecuadamente por una línea recta.

Coeficiente de determinación ajustado, R^2_{aj}. Este coeficiente se calcula de la siguiente manera:

$$R^2_{aj} = \frac{CM_{\text{Total}} - CM_E}{CM_{\text{Total}}}$$

donde el cuadrado medio total, CM_{Total}, se obtiene al dividir la suma de cuadrados total, S_{yy}, entre sus grados de libertad. Cuando hay muchos términos en un modelo, el estadístico R^2_{aj} se prefiere en lugar de R^2, puesto que este último es engañoso al incrementarse en forma artificial con cada término que se agrega al modelo, aunque sea un término que no contribuya en nada a la explicación de la respuesta. En cambio, el R^2_{aj} incluso baja de valor cuando el término que se agrega no aporta nada. Se cumple que $0 < R^2_{aj} \leq R^2 \leq 1.0$. En general, para fines de predicción se recomienda un coeficiente de determinación ajustado de al menos 0.7.

Para el modelo del ejemplo 11.1, de acuerdo a la tabla 11.7, el coeficiente de determinación ajustado está dado por:

$$R^2_{aj} = \frac{[(2\,580.86)/(13)] - [(180.32)/12]}{(2\,580.86)/(13)} = 0.92431$$

Coeficiente de correlación
Mide la intensidad de la relación lineal entre dos variables.

Coeficiente de correlación. Es bien conocido que el *coeficiente de correlación*, r, mide la intensidad de la relación lineal entre dos variables X y Y. Si se tiene n pares de datos de la forma (x_i, y_i), entonces este coeficiente se obtiene de la siguiente manera:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

Se puede ver que $-1 \leq r \leq 1$; si r es próximo a -1, entonces tendremos una relación lineal negativa fuerte, y si r es próximo a cero, entonces diremos que no hay correlación lineal, y finalmente si r es próximo a 1, entonces tendremos una relación lineal positiva fuerte. Por ejemplo, para los datos del ejemplo 11.1, el coeficiente de correlación es $r = (1\,478)/\sqrt{(910)(2\,580.9)} = 0.96442$, lo cual habla de una correlación lineal positiva fuerte.

Es importante notar que sólo en el caso particular del modelo de la línea recta de regresión, existe una relación directa entre r y R^2, ya que:

$$R^2 = \frac{SC_R}{S_{yy}} = \frac{\hat{\beta}_1 S_{xy}}{S_{yy}} = \frac{\frac{S_{xy}}{S_{xx}} S_{xy}}{S_{yy}} = \frac{(S_{xy})^2}{S_{xx}S_{yy}} = r^2$$

Error estándar de estimación
Estima la desviación estándar del error e indica la magnitud del error de estimación de un modelo.

Error estándar de estimación. Una medición sobre la calidad del ajuste de un modelo lo da el *error estándar de estimación*, que es una estimación de la desviación estándar del error, σ. En el caso de la regresión lineal simple, está dado por:

$$\hat{\sigma} = \sqrt{\frac{SC_E}{n-2}} = \sqrt{\frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2}} \qquad \textbf{(11.36)}$$

Es claro que a medida que el modelo ajuste mejor, la SC_E será menor y en consecuencia el error estándar de estimación también será menor. En los datos del ejemplo 11.1, $\hat{\sigma} = \sqrt{15.0271} = 3.8765$.

Media del error absoluto (*mea*)
Es la media del valor absoluto de los residuos, y sirve para ver cuánto falla en promedio el modelo al hacer la estimación de la variable de respuesta.

Media del error absoluto (*mea*). Otra forma de medir la calidad del ajuste es a través de la media del valor absoluto de los residuos, es decir:

$$mea = \frac{\Sigma_{i=1}^{n} |e_i|}{n} \qquad \textbf{(11.37)}$$

es claro que mientras mejor sea el ajuste, los residuos serán más pequeños y, en consecuencia, también la *mea* tenderá a ser más pequeña. La *mea* se puede ver como una medición para ver cuánto falla en promedio el modelo al hacer la estimación de la variable de respuesta. En los datos del ejemplo 11.1, de acuerdo a la tabla 11.3:

$$mea = \frac{|-3.2|+|4.6|+|-1.7|+\ldots+|3.6|}{14} = \frac{43.37}{14} = 3.0979$$

Análisis gráfico de residuos

Como complemento a lo que se ha discutido hasta aquí, un análisis adecuado de los residuos proporciona información adicional sobre la calidad del ajuste del modelo de regresión y de esa manera es posible verificar si el modelo es adecuado. Las gráficas que suelen hacerse para completar el diagnóstico del modelo consisten en: graficar los residuos en papel de probabilidad normal, graficar los residuos contra los predichos, los residuos contra cada variable regresora y contra alguna otra variable importante que no haya sido incluida en el modelo.

Por ejemplo, para los residuos del ejemplo 11.1 que se muestran en la tabla 11.3, se construye la gráfica de probabilidad normal que se muestra en la figura 11.3. En ésta se aprecia que el supuesto de normalidad sobre los errores se cumple razonablemente bien, ya que los puntos en esta gráfica tienden a ajustarse a la línea recta.

A partir de la tabla 11.3 es fácil obtener la gráfica de residuos contra predichos (e_i *vs.* $\hat{y}_i$) que se muestra en la figura 11.4. Si el modelo es adecuado se espera que en esta gráfica los puntos no sigan ningún patrón y que, por lo tanto, estén distribuidos más o menos aleatoriamente a lo largo y ancho de la gráfica. Cuando esto ocurre significa que el modelo se ajusta de igual manera a lo largo de los valores de *Y*. Por el contrario, si se aprecia algún patrón habrá que ver cuál es el tipo de patrón que se observa en la gráfica y diagnosticar cuál es la falla que registra el modelo.

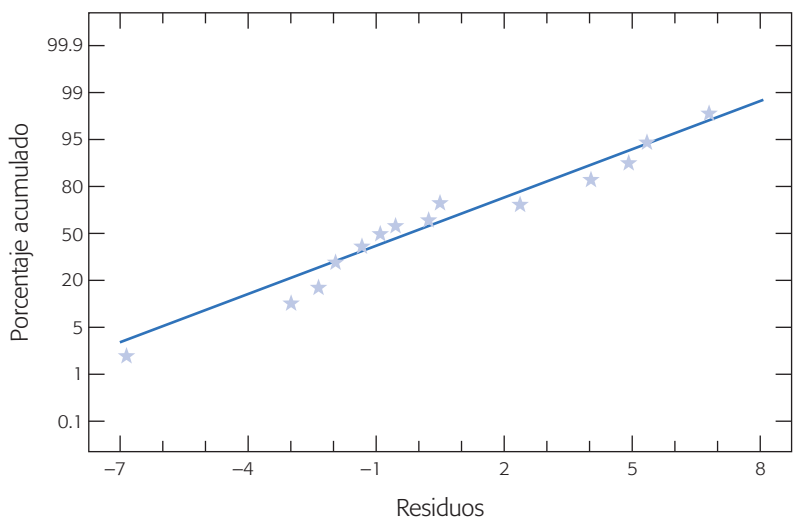


Figura 11.3 Gráfica de probabilidad normal para los residuos del ejemplo 11.1.

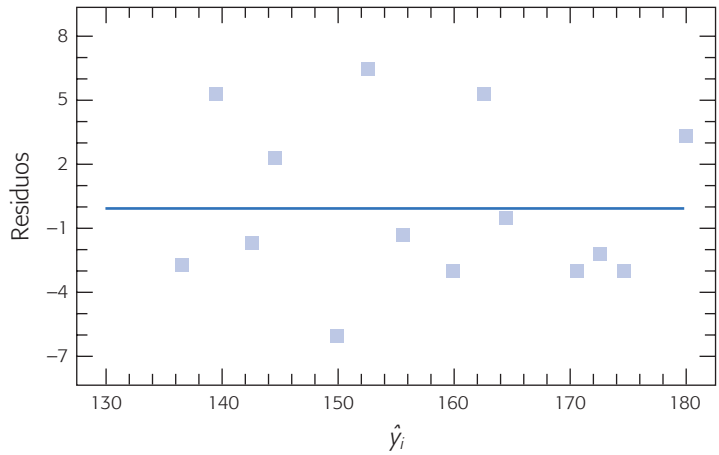


Figura 11.4 Gráfica de residuos contra estimados o predichos, $\hat{y}_i$, para el ejemplo 11.1.

En particular la figura 11.4 no muestra ninguna anomalía, lo cual es una evidencia más a favor del modelo de regresión simple para el ejemplo 11.1.

En la figura 11.5 se muestra la gráfica de los residuos contra los valores de X, para el ejemplo 11.1. Esa gráfica se interpreta de igual manera que la anterior, y en ella tampoco se observa alguna situación anormal.

Lo comentado en el capítulo 3 sobre las gráficas de residuos y su interpretación también aplica a los residuos de un modelo de regresión, por lo que remitimos al lector a tal capítulo si desea una discusión e interpretación más detallada de las gráficas anteriores.

Verificación del supuesto de independencia

Uno de los supuestos importantes sobre los residuos es que éstos son independientes. Para verificar tal suposición se suelen graficar los residuos contra el orden en el que se obtuvieron los datos. La suposición se cumple si los puntos en esta gráfica no siguen ningún patrón bien definido, como alguna tendencia (véase por ejemplo la figura 3.5*f* del capítulo 3).

Prueba de Durbin-Watson
Diagnostica la presencia de correlación entre los residuos consecutivos, que es una posible manifestación de la falta de independencia.

Prueba de Durbin-Watson. Esta prueba permite diagnosticar la presencia de correlación (autocorrelación) entre los residuos consecutivos (ordenados en el tiempo), que es una posible manifestación de la falta de independencia. La *autocorrelación* se presenta en experimentos en los cuales cada medición tiene alguna contaminación de la medición inmediata anterior, lo cual contradice el supuesto de independencia. Sea ρ el parámetro que representa la correlación entre residuos consecutivos ($Corr\ (e_t, e_{t-1}) = \rho$; $t = 2, 3, ..., n$). La hipótesis en la *prueba de Durbin-Watson* es:

$$\begin{aligned} H_0 &: \rho = 0 \\ H_A &: \rho > 0 \end{aligned} \qquad \textbf{(11.38)}$$

donde la alternativa se toma en el sentido mayor (>) porque la autocorrelación positiva es la más frecuente en la práctica. En la gráfica de residuos contra el tiempo se observa autocorrelación positiva cuando los puntos caen encadenados (como en la

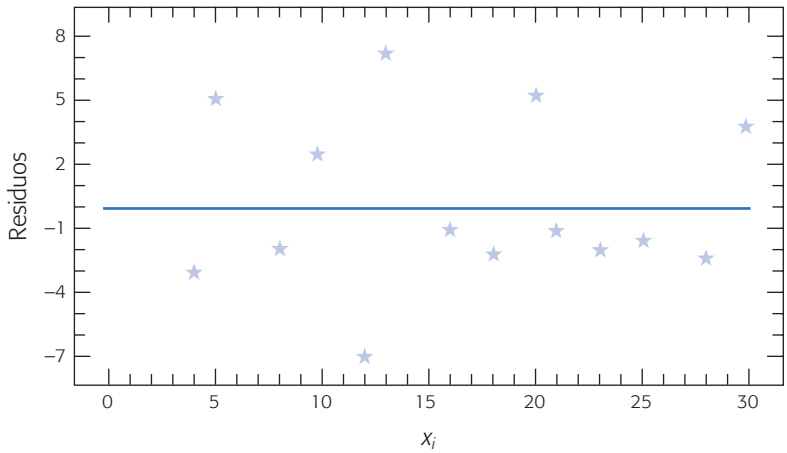


Figura 11.5 Gráfica de residuos contra x_i, para el ejemplo 11.1.

figura 3.5*f*). Por el contrario, cuando los puntos se van alternando de negativos a positivos la autocorrelación es negativa. El estadístico de Durbin-Watson es:

$$d = \frac{\Sigma_{i=2}^{n}(e_i - e_{i-1})^2}{\Sigma_{i=1}^{n}(e_i)^2} \tag{11.39}$$

donde los e_i, $i = 1, 2, ..., n$, son los residuos ordenados en el tiempo. La decisión sobre la hipótesis dada en (11.38), consiste en la siguiente regla:

$$\begin{array}{ll} \text{Si } d < d_L & \text{Se rechaza } H_0 \\ \text{Si } d > d_U & \text{No se rechaza } H_0 \\ \text{Si } d_L \leq d \leq d_U & \text{Sin decisión} \end{array} \tag{11.40}$$

donde d_L y d_U son cotas que se leen en tablas dadas en el apéndice. Para entrar a las tablas se requiere el número de residuos n, el nivel de significancia prefijado α y el número de variables explicativas o regresoras en el modelo, p.

En caso de interesar la hipótesis de autocorrelación negativa ($H_A : p < 0$) se utiliza el estadístico $d' = 4 - d$, donde d se calcula con la ecuación (11.39). En caso de interesar la hipótesis bilateral con alternativa $H_A : p \neq 0$, se combinan las dos pruebas unilaterales de tamaño α de manera que la prueba bilateral tenga el tamaño deseado 2α.

La prueba de Durbin-Watson tiene el inconveniente de detectar sólo la estructura de correlación de residuos consecutivos. No detecta correlaciones entre residuos no consecutivos en el tiempo que también violan el supuesto de independencia. Este tipo de correlación ocurre en un experimento cuando la contaminación de una medición a otra no se refleja de inmediato, sino que actúa con retardo.

Prueba de falta de ajuste

Prueba de falta de ajuste
Verifica la calidad del ajuste del modelo, en particular ayuda a evaluar si el orden del modelo es el correcto.

Aunque con las pruebas que hemos visto hasta aquí suele ser suficiente para evaluar la calidad del ajuste del modelo, cuando por al menos un valor x hay varias observaciones de Y, es posible desarrollar una prueba adicional, que se conoce como *prueba de falta de ajuste*. Ésta es útil cuando se quiere verificar si el orden del modelo es

correcto, por ejemplo, cuando se suponga un efecto lineal y en realidad hay una curvatura (ver capítulo 12). La hipótesis que se va a probar con esta prueba es:

H_0 : El modelo se ajusta de manera adecuada a los datos
H_A : El modelo no se ajusta en forma satisfactoria

Cuando un modelo no se ajusta bien a los datos, los residuos estarán influidos no sólo por los errores experimentales o error aleatorio, sino también por el error por falta de ajuste. La idea de esta prueba es separar estos dos tipos de errores, por ello la suma de cuadrados del error se divide en dos componentes:

$$SC_E = SC_{EP} + SC_{FA}$$

donde SC_{EP} es la suma de cuadrados atribuibles a un error experimental "puro" y SC_{FA} es la suma de cuadrados atribuible a la falta de ajuste del modelo. Para estimar SC_{FA} será necesario que para al menos un valor x_i haya varias observaciones de Y. Supongamos que las n observaciones se pueden arreglar de la siguiente manera:

$Y_{ll}, Y_{12},\ldots Y_{1n_1}$ son las n_1 observaciones repetidas para X_1; y sea $\overline{y}_1$ la media de éstas.

$Y_{21}, Y_{22},..., Y_{2n_2}$ son las n_2 observaciones repetidas para x_2; y sea $\overline{y}_2$ la media de éstas.

$\vdots$

$Y_{m1}, Y_{m2},..., Y_{mn_m}$ son las n_m observaciones repetidas para x_m; y sea $\overline{y}_m$ la media de éstas.

Se tienen en total m niveles distintos de X, y además se cumple que $\Sigma_{i=1}^{m} n_i = n$. Es claro que la diferencia entre los valores de Y dentro de cada nivel x_i se debe a un error experimental no atribuible; por ello, la suma del cuadrado del error puro se estima de la siguiente manera:

$$SC_{EP} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{m} n_i (\overline{y}_i)^2$$

En total, esta suma de cuadrados tiene $\Sigma_{i=1}^{m} (n_i - 1) = n - m$ grados de libertad. La suma de cuadrados por falta de ajuste se obtiene con:

$$SC_{FA} = SC_E - SC_{EP}$$

la cual tiene $m - 2$ grados de libertad. Con esto, el estadístico para probar la falta de ajuste está dado por:

$$F_0 = \frac{SC_{FA}/(m-2)}{SC_{EP}/(n-m)} = \frac{CM_{FA}}{CM_{EP}}$$

Bajo H_0 este estadístico tiene una distribución F con $(m-2)$ y $(n-m)$ grados de libertad en el numerador y denominador, respectivamente. Por lo tanto, se rechaza H_0 si $F_0 > F_{(\alpha, m-2, n-m)}$; o en forma equivalente, si valor-$p = P(F > F_0) < \alpha$.

El resultado de esta hipótesis debe ser comparado con el resto de los criterios sobre calidad de los modelos vistos antes. Si están orientados en el mismo sentido, entonces la prueba de falta de ajuste será un elemento adicional a favor o en contra del modelo.

Los *buenos modelos* son aquellos que cumplen más criterios de calidad del ajuste. Siempre existirán circunstancias en las que, al no cumplirse alguno de los criterios, desde el punto de vista práctico no necesariamente harán inviable el modelo. Por ejemplo, si la normalidad en los residuos no se cumple, se sabe que esa suposición no es tan fuerte, es decir, que la metodología es más o menos robusta a la falta de normalidad. Otro aspecto a tomar en cuenta es que bajo calidad similar en el ajuste de dos modelos, siempre se deberá preferir el más sencillo.

Estimación y predicción por intervalo en regresión simple

Una de las aplicaciones más importantes en un análisis de regresión es hacer estimaciones de la respuesta media para un valor dado x_0. En el caso particular de la regresión lineal simple, sabemos que un estimador puntual de la respuesta media lo da la recta de regresión:

$$\hat{E}(y|x_0) \equiv \hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

Además de esto, en ocasiones es de interés obtener una estimación por intervalo para $E(y \mid x_0)$. Para ello, como

$$V(\hat{y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

al sustituir σ^2 por su estimador CM_E, y como $\hat{y}_0$ hereda la distribución normal de $\hat{\beta}_0$ y $\hat{\beta}_1$, entonces un intervalo de confianza al $100(1-\alpha)\%$ para la respuesta media en x_0 está dado por:

$$\hat{y}_0 - t_{(\alpha/2, n-2)} \sqrt{CM_E \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]} \leq E(y|x_0) \leq \hat{y}_0 + t_{(\alpha/2, n-2)} \sqrt{CM_E \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]} \quad \textbf{(11.41)}$$

A este intervalo se le conoce como *intervalo para la recta de regresión*. Note que su amplitud depende del CM_E y de la distancia entre x_0 y $\bar{x}$. La amplitud es mínima cuando $x_0 = \bar{x}$ y se incrementa conforme $|x_0 - \bar{x}|$ se hace más grande.

Para ilustrar lo anterior consideremos el modelo ajustado a los datos del ejemplo 11.1, y obtengamos el intervalo de confianza para la respuesta media en $x_0 = 12$ (porcentaje de fibra).

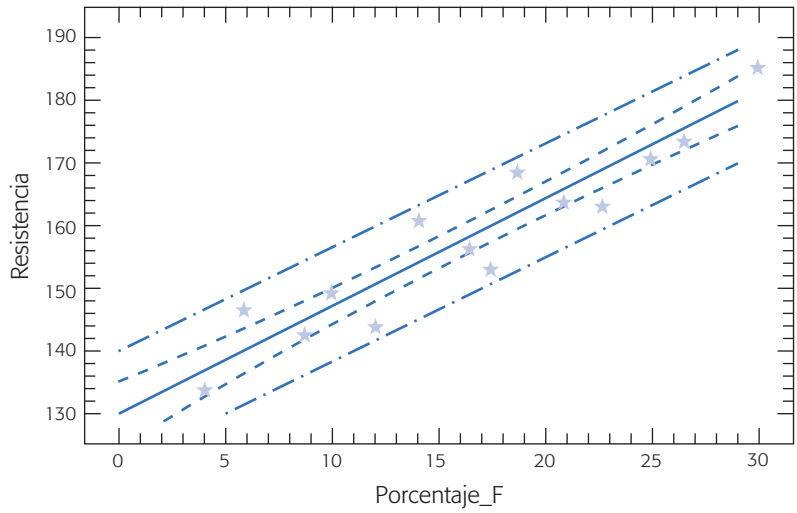


Figura 11.6 Recta de regresión con intervalo de confianza y de predicción para observaciones futuras para el ejemplo 11.1.

El estimador puntual está dado por $\hat{y} = 130.67 + (1.6242)(12) = 150.16$; y un intervalo de confianza al 95% para la respuesta media en ese punto es:

$$150.16 \pm 2.179\sqrt{15.0271\left[\frac{1}{14}+\frac{(12-17)^2}{910}\right]}$$

$$150.16 \pm 2.6564$$

De aquí que el intervalo de confianza para la respuesta media en $x_0 = 12$ está dado por:

$$147.5 \leq E(y|x_0 = 12) \leq 152.82$$

La primera banda de confianza en torno a la recta de regresión de la figura 11.6 indica el cálculo del intervalo de confianza de la expresión (11.41) para los valores x_0 con los que se ajustó la recta de regresión. Nótese cómo se va abriendo esta banda debido a que se incrementa $|x_0 - \bar{x}|$.

Predicción de observaciones futuras

Una de las aplicaciones más útiles de un modelo de regresión es predecir o pronosticar nuevas o futuras observaciones de Y. Una estimación puntual de la observación futura de y_0 en el punto x_0 está dada por:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \qquad \textbf{(11.42)}$$

Para predecir por intervalo, como la nueva observación es independiente de las observaciones utilizadas para ajustar el modelo de regresión, el intervalo para la recta de regresión dado por la expresión (11.41) no es apropiado. Por ello, sea y_0 la

observación futura en x_0 y sea $\hat{y}_0$ dado por la ecuación (11.42) el estimador puntual para y_0; de aquí definimos el error de esta estimación como:

$$\varepsilon_0 = y_0 - \hat{y}_0$$

esta variable, ε_0, se distribuye de manera normal con media cero y varianza,

$$\begin{aligned} V(\varepsilon_0) &= V(y_0 - \hat{y}_0) = V(y_0) + V(\hat{y}_0) \\ &= \sigma^2 + \sigma^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \\ &= \sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \end{aligned}$$

Dada esta varianza para el error de predicción, y estimando σ^2 con el CM_E, entonces el intervalo de predicción al $100(1 - \alpha)\%$ con respecto a la observación futura en x_0 está dado por:

$$\hat{y}_0 - t_{(\alpha/2,\, n-2)}\sqrt{CM_E\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]} \le y_0 \le \hat{y}_0 + t_{(\alpha/2,\, n-2)}\sqrt{CM_E\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]} \quad \textbf{(11.43)}$$

También este intervalo de predicción se amplía a medida que aumenta $|x_0 - \bar{x}|$. Al comparar las expresiones (11.41 y 11.43), se aprecia que el intervalo para la predicción siempre es más amplio que el intervalo de confianza en x_0. Esto se debe a que, como ya lo vimos, el intervalo de predicción depende tanto del error del modelo ajustado como del error asociado a las observaciones futuras (σ).

De igual forma y siguiendo el mismo razonamiento, es posible encontrar un intervalo de predicción del $100(1 - \alpha)\%$ para la media de k observaciones futuras en $X = x_0$. Sea $\bar{y}_0$ la media de k observaciones futuras en $X = x_0$, entonces el intervalo de predicción del $100(1 - \alpha)\%$ para $\bar{y}_0$ está dado por:

$$\begin{aligned} &\hat{y}_0 - t_{(\alpha/2,\, n-2)}\sqrt{CM_E\left[\frac{1}{k} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]} \\ &\le \bar{y}_0 \le \bar{y}_0 + t_{(\alpha/2,\, n-2)}\sqrt{CM_E\left[\frac{1}{k} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]} \end{aligned}$$

Para ilustrar la aplicación de la predicción por intervalo, supongamos que deseamos predecir por intervalo cuál es la resistencia esperada en un nuevo experimento cuando se añada $x_0 = 12\%$ de fibra. Con la ecuación (11.43) se obtiene que el intervalo de predicción es:

$$\begin{aligned} &150.16 - 2.179\sqrt{15.0271\left[1 + \frac{1}{14} + \frac{(12-17)^2}{910}\right]} \\ &\le y_0 \le 150.16 + 2.179\sqrt{15.0271\left[1 + \frac{1}{14} + \frac{(12-17)^2}{910}\right]} \end{aligned}$$

que al simplificar queda como $141.31 \le y_0 \le 159.01$.

La segunda banda en torno a la recta de regresión de la figura 11.6 indica el cálculo del intervalo de predicción dado por la expresión (11.43). Nótese cómo se va abriendo esta banda debido a que se incrementa $|x_0 - \bar{x}|$.

Riesgos de la regresión

Un primer riesgo del análisis de regresión es que, a partir de un modelo significativo, siempre se concluya de manera directa una relación causa-efecto entre X y Y. En ocasiones, esta conclusión puede ser falsa, ya que al estar relacionadas dos variables no necesariamente implica que hay una relación causa-efecto. Estrictamente hablando, lo único que indica que un análisis de regresión que es significativo es que existe la relación que respalda el modelo, y el usuario es quien debe investigar si tal relación es de tipo causa-efecto. Esto puede ser más o menos difícil dependiendo del origen de los datos. Recordemos que al inicio de este capítulo se dijo que los datos para hacer un análisis de regresión pueden originarse de experimentos planeados, de observaciones de fenómenos no controlados o de registros históricos.

En cualquier interpretación de las razones de una relación significativa se debe recurrir al conocimiento del proceso. Además, se debe tomar en cuenta que algunas de las razones por las que las variables X y Y aparecen relacionadas de manera significativa son:

- X influye sobre Y.
- Y influye sobre X.
- X y Y interactúan entre sí, una tercera variable Z influye sobre ambas y es la causante de tal relación.
- X y Y actúan en forma similar debido al azar.
- X y Y aparecen relacionados debido a que la muestra no es representativa.

Otro riesgo es hacer extrapolaciones indiscriminadas con base en el modelo. Para no incurrir en esto cuando se quieran predecir nuevas observaciones o estimar la respuesta media en algún punto x_0, se debe tener cuidado en cuanto a extrapolar más allá de la región que contienen las observaciones originales. Es probable que un modelo que ajusta bien en la región de los datos originales ya no ajustará bien fuera de esa región. Esto se debe a que quizá muy fuera de la región de los datos originales empiecen a actuar otros fenómenos no considerados en el modelo original. Este riesgo es más grande en el análisis de regresión múltiple, ya que se trabaja con regiones multidimensionales.

Regresión lineal múltiple

En muchas situaciones prácticas existen varias variables independientes que se cree que influyen o están relacionadas con una variable de respuesta Y, y por lo tanto será necesario tomar en cuenta si se quiere predecir o entender mejor el comportamiento de Y. Por ejemplo, para explicar o predecir el consumo de electricidad en una casa habitación tal vez sea necesario considerar el tipo de residencia, el número de personas que la habitan, la temperatura promedio de la zona, etcétera.

Sea $X_1, X_2, \ldots, X_k$ variables independientes o regresoras, y sea Y una variable de respuesta, entonces el *modelo de regresión lineal múltiple* con k variables independientes es el polinomio de primer orden:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \qquad \textbf{(11.44)}$$

donde los β_j son los parámetros del modelo que se conocen como *coeficientes de regresión* y ε es el error aleatorio, con media cero, $E(\varepsilon) = 0$ y $V(\varepsilon) = \sigma^2$. Si en la ecuación (11.44) $k = 1$, estamos en el caso de regresión lineal simple y el modelo es una línea recta; si $k = 2$, tal ecuación representa un plano. En general, la ecuación (11.44) representa un hiperplano en el espacio de k dimensiones generado por las variables $\{X_j\}$.

El término lineal del modelo de regresión se emplea debido a que la ecuación (11.44) es función lineal de los parámetros desconocidos $\beta_0, \beta_1, \ldots, \beta_k$. La interpretación de éstos es muy similar a lo ya explicado para el caso de regresión lineal simple: β_0 es la ordenada al origen, y β_j mide el cambio esperado en Y por cambio unitario en X_j cuando el resto de las variables regresoras se mantienen fijas o constantes.

Es frecuente que en la práctica se requieran modelos de mayor orden para explicar el comportamiento de Y en función de las variables regresoras. Por ejemplo, supongamos que se tienen dos variables independientes y que se sospecha que la relación entre Y y algunas de las variables independientes es cuadrática, por ello quizá se requiera un polinomio de segundo orden como modelo de regresión:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \varepsilon \qquad \textbf{(11.45)}$$

Éste también es un modelo de regresión lineal múltiple, ya que la ecuación (11.45) es una función lineal de los parámetros desconocidos $\beta_0, \beta_1, \ldots, \beta_{22}$. Pero además si definimos $X_3 = X_1 X_2$, $\beta_3 = \beta_{12}$, $X_4 = X_1^2$, $\beta_4 = \beta_{11}$, $X_5 = X_2^2$ y $\beta_5 = \beta_{22}$; entonces, la ecuación (11.45) puede escribirse así:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

la cual tiene la misma forma que el modelo general de regresión lineal múltiple de la expresión (11.44). Con lo visto antes, estamos en posibilidades de abordar el problema de estimación de los parámetros del modelo de regresión múltiple, que será aplicable a una amplia gama de modelos que pueden reducirse a la forma general de la expresión (11.44).

Para estimar los parámetros de la regresión lineal múltiple se necesita contar con n datos ($n > k$), que tienen la estructura descrita en la tabla 11.8. En ésta se aprecia que para cada combinación de valores de las variables regresoras, $(x_{1i}, \ldots, x_{ki})$ se observa un valor de la variable dependiente, y_i.

En términos de los datos, el modelo de regresión lineal múltiple puede escribirse de la siguiente manera:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i \\ &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ji} + \varepsilon_i, \qquad i = 1, 2, \ldots, n \end{aligned} \qquad \textbf{(11.46)}$$

Tabla 11.8 Estructura de los datos para la regresión lineal múltiple.

Y	X_1	X_2	...	X_k
y_1	x_{11}	x_{21}	...	x_{k1}
y_2	x_{12}	x_{22}	...	x_{k2}
	⋮	⋮		⋮
y_n	y_{1n}	x_{2n}	...	x_{kn}

Al despejar los errores, elevarlos al cuadrado y sumarlos obtenemos la siguiente función:

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ji} \right)^2 \qquad \textbf{(11.47)}$$

esta función depende de los parámetros β_j. Los estimadores de mínimos cuadrados para β_j se obtienen al minimizar los errores, es decir, minimizando S. Esto se logra si derivamos a S con respecto a cada parámetro β_j, $\frac{\partial S}{\partial \beta_j}$, $(j = 0, 1, 2, \ldots, k)$, las $k + 1$ ecuaciones resultantes se igualan a cero. La solución de las $k + 1$ ecuaciones simultáneas son los estimadores de mínimos cuadrados, $\hat{\beta}_j$.

Ilustrar el procedimiento de estimación por mínimos cuadrados es más sencillo si se utiliza notación matricial. En términos de los datos, ecuación (11.46), el modelo puede escribirse en notación matricial como

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

donde,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \ldots & x_{k1} \\ 1 & x_{12} & x_{22} & \ldots & x_{k2} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{kn} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{y} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Queremos encontrar el vector de los estimadores de mínimos cuadrados, $\hat{\beta}$, que minimice

$$\begin{aligned} S &= \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) \\ &= \mathbf{y}'\mathbf{y} - \beta'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta \\ &= \mathbf{y}'\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{y} + \beta'\mathbf{X}'\mathbf{X}\beta \end{aligned}$$

La última igualdad se debe a que $\beta'\mathbf{X}'\mathbf{y}$ es una matriz (1×1), o un escalar, y por lo tanto su transpuesta $(\beta'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\beta$ es el mismo escalar. De aquí que los estimadores de mínimos cuadrados deban satisfacer la siguiente expresión:

$$\left.\frac{\partial S}{\partial \beta}\right|_{\hat{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = 0$$

esto implica que:

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y} \tag{11.48}$$

Para resolver esta ecuación en términos de $\hat{\beta}$, se multiplica por ambos lados de la ecuación (11.48) por la matriz inversa de $\mathbf{X}'\mathbf{X}$, y se obtiene que el estimador de mínimos cuadrados de β es:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{11.49}$$

por lo tanto, el modelo ajustado está dado por:

$$\hat{y} = \mathbf{X}\hat{\beta} \tag{11.50}$$

Además, se puede demostrar que $\hat{\beta}$ es un estimador insesgado, $E(\hat{\beta}) = \beta$, y la matriz de covarianza de $\hat{\beta}$ es:

$$Cov\,(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Para hacer inferencias sobre β o, en general sobre el modelo, es necesario encontrar una forma de estimar σ^2. A partir de la ecuación (11.50) es claro que el vector de residuos está dado por $\mathbf{e} = \mathbf{y} - \hat{y} = \mathbf{y} - \mathbf{X}\hat{\beta}$. De aquí que la suma de cuadrados del error esté dada por:

$$\begin{aligned} SC_E &= \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} \\ &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \end{aligned} \tag{11.51}$$

De acuerdo con la ecuación (11.48): $\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{y}$, entonces esta última ecuación toma la siguiente forma:

$$SC_E = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} \tag{11.52}$$

La suma de cuadrados del error dada por esta última expresión tiene $n - k - 1$ grados de libertad, donde $k + 1$ es igual al número de parámetros estimados en el modelo. Entonces, el *cuadrado medio del error* es:

$$CM_E = \frac{SC_E}{n - k - 1}$$

Se puede demostrar que el valor esperado de CM_E es σ^2, por lo que es natural que el estimador de σ^2 esté dado por:

$$\hat{\sigma}^2 = CM_E \tag{11.53}$$

La raíz cuadrada del CM_E se conoce como *error estándar de estimación* del modelo.

Ejemplo 11.3

En Ramírez *et al*. (2001) se presenta un experimento secuencial para optimizar la producción de un colorante natural. En la etapa final se delimitó una zona de experimentación donde se sospecha que se encuentran las condiciones óptimas para la producción de este colorante en función de la concentración de carbono (X_1) y temperatura (X_2). En la tabla 11.9 se muestran los niveles de X_1 y X_2 con los que se experimentó, así como la producción observada en cada una de las condiciones.

A continuación ajustaremos un modelo de segundo orden:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \varepsilon_i$$

De aquí que si expresamos esto en forma matricial, $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, toma la siguiente forma (sólo se muestra parcialmente):

$$\begin{matrix} & 1 & x_1 & x_2 & x_1x_2 & x_1^2 & x_2^2 \end{matrix}$$

$$\begin{bmatrix} 5\,707 \\ 5\,940 \\ 3\,015 \\ \vdots \\ 7\,543 \end{bmatrix} = \begin{bmatrix} 1 & 9 & 17 & 153 & 81 & 289 \\ 1 & 13 & 17 & 221 & 169 & 289 \\ 1 & 9 & 25 & 225 & 81 & 625 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 11 & 21 & 231 & 121 & 141 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{12} \\ \beta_{11} \\ \beta_{22} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \vdots \\ \varepsilon_{12} \end{bmatrix}$$

A partir de aquí se obtiene $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, que al hacer los cálculos obtenemos el siguiente modelo ajustado:

$$Y = -75\,732.8 + 4\,438.69X_1 + 5\,957.79X_2 - 17.9688X_1X_2 - 181.316X_1^2 - 146.404X_2^2$$

Tabla 11.9 Datos para el ejemplo 11.3.

X_1: Carbono	X_2: Temperatura	Y: Producción
9	17	5 707
13	17	5 940
9	25	3 015
13	25	2 673
8.17	21	5 804
13.8	21	6 700
11	15.34	5 310
11	26.66	725
11	21	7 521
11	21	7 642
11	21	7 500
11	21	7 545

En la tabla 11.10 se presentan las observaciones, los valores predichos y los residuos de este modelo. En la figura 11.7 se muestran gráficas de estos residuos para diagnosticar la calidad de ajuste del modelo. La interpretación de estas gráficas es similar a lo explicado para la regresión lineal simple. En la figura 11.7*a*) se muestra la gráfica de probabilidad normal para los residuos, en ésta se aprecia que la normalidad se cumple de manera satisfactoria.

La figura 11.7*b*) corresponde a la gráfica de residuos contra predichos, donde se observa una ligera tendencia en forma de embudo que podría indicar un mayor error de ajuste para valores grandes de la variable de respuesta; pero, al observar con detenimiento la distribución de los puntos se aprecia que la apariencia referida básicamente se debe a dos puntos (los residuos 2 y 6). Por ello, de acuerdo con los resultados de esta gráfica podemos considerar que el supuesto de varianza constante se cumple aceptablemente.

En la figura 11.7*c*) se muestran los residuales contra los niveles de temperatura y no se nota ningún patrón fuerte. En la figura 11.7*d*) se aprecia la gráfica de residuales contra los valores de carbono, y en ésta se observa una ligera tendencia de embudo, pero no demasiado fuerte; por ello, de acuerdo con este criterio podemos considerar que el modelo es aceptable.

Pruebas de hipótesis en regresión lineal múltiple

Las hipótesis sobre los parámetros del modelo son equivalentes a las realizadas para regresión lineal simple, pero ahora son más necesarias porque en regresión múltiple tenemos más parámetros en el modelo; sin embargo, por lo general es necesario evaluar su verdadera contribución a la explicación de la respuesta. También requerimos de la suposición de que los errores se distribuyen en forma normal, independientes, con media cero y varianza σ^2 ($\varepsilon_i \sim \text{NID}(0, \sigma^2)$). Una consecuencia de esta suposición es que las observaciones y_i son: $\text{NID}(\beta_0 + \Sigma_{j=1}^{k} \beta_j X_{ji}, \sigma^2)$.

Tabla 11.10 Valores observados predichos y residuos para el ejemplo 11.3.

y_j	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$
5 707	5 751.1	–44.1
5 940	6 328.2	–388.2
3 015	2 927.9	87.1
2 673	2 929.9	–256.9
5 804	5 896.7	–92.7
6 700	6 306.2	393.8
5 310	5 066.7	243.3
725	667.3	57.7
7 521	7 552.0	–31.0
7 642	7 552.0	90.0
7 500	7 552.0	–52.0
7 545	7 552.0	–7.0

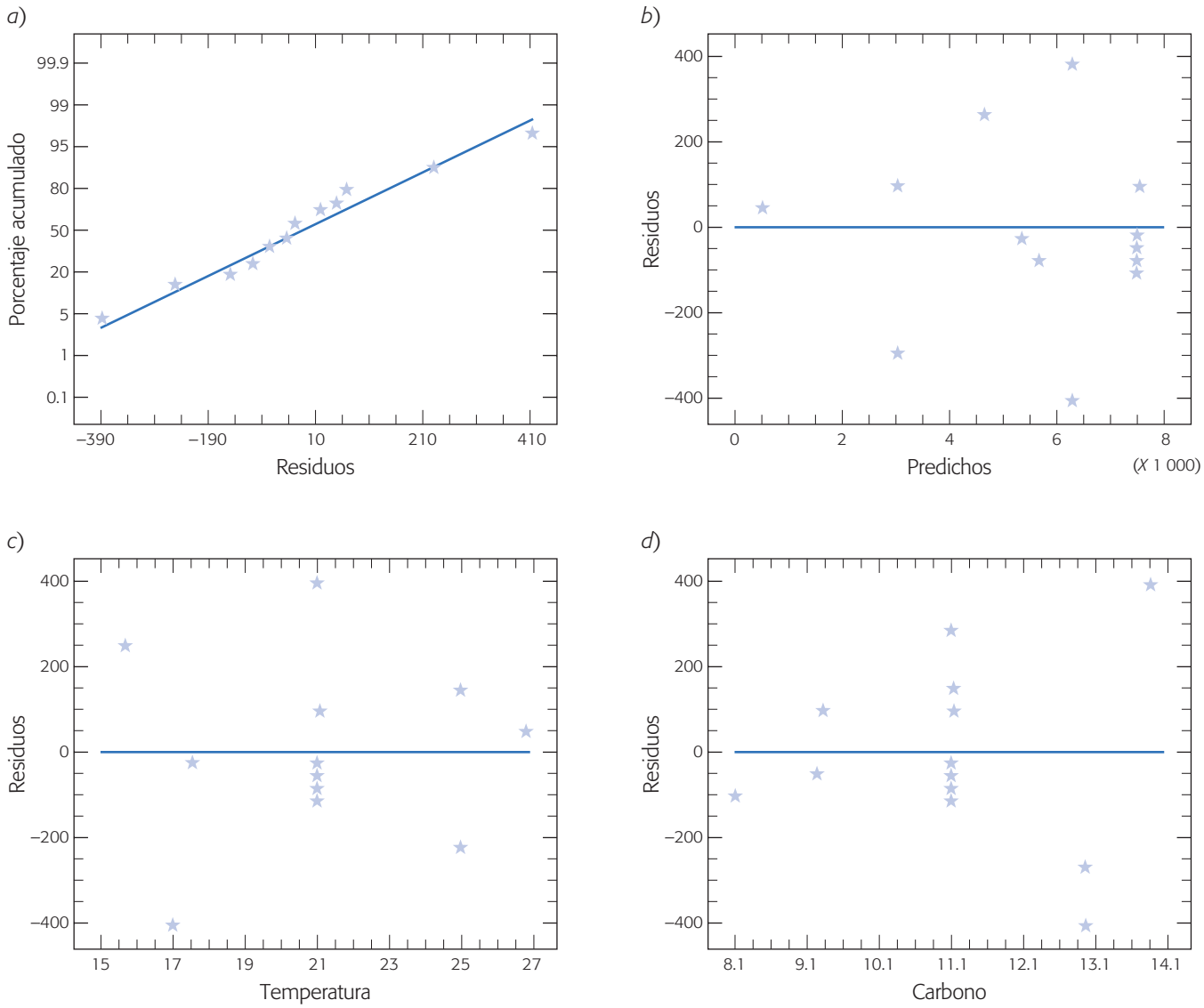


Figura 11.7 Gráficas de residuos para el ejemplo 11.3, *a*) probabilidad normal, *b*) residuales contra predichos, *c*) residuales contra los niveles de temperatura, *d*) residuales contra los niveles de carbono.

Análisis de varianza

La hipótesis global más importante sobre un modelo de regresión múltiple consiste en ver si la regresión es significativa. Esto se logra probando la siguiente hipótesis:

$$H_0 : \beta_1 = \beta_2 = \ldots \beta_k = 0$$
$$H_A : \beta_j \neq 0 \qquad \text{para al menos un } j = 1, 2, \ldots, k$$

Aceptar H_0 significa que ningún término o variable en el modelo tiene una contribución significativa al explicar la variable de respuesta, Y. Mientras que rechazar

H_0 implica que por lo menos un término en el modelo contribuye de manera significativa a explicar Y. El procedimiento para probar esta hipótesis es una generalización del procedimiento utilizado para probar la hipótesis equivalente, expresión (11.32), en regresión lineal simple. Al igual que la expresión (11.30), aquí también se descompone la suma total de cuadrados en la suma de cuadrados de regresión y en la suma de cuadrados del error:

$$S_{yy} = SC_R + SC_E \tag{11.54}$$

Si $H_0 : \beta_j = 0$ es verdadera, entonces SC_R/σ^2 tiene una distribución χ^2_k, donde el número de grados de libertad, k, es igual al número de términos en el modelo de regresión. Además, $SC_E/\sigma^2 \sim \chi^2_{n-k-1}$, y SC_E y SC_R son independientes. Luego, es natural que el estadístico de prueba para la significancia del modelo de regresión lineal múltiple esté dado por:

$$F_0 = \frac{SC_R/k}{SC_E/(n-k-1)} = \frac{CM_R}{CM_E} \tag{11.55}$$

que bajo H_0 tiene una distribución $F_{(k,\, n-k-1)}$. Así, se rechaza H_0 si $F_0 > F_{(\alpha,\, k,\, n-k-1)}$ o también si valor-$p = P(F > F_0) < \alpha$.

Para completar el procedimiento anterior necesitamos una forma explícita para calcular SC_R. En la ecuación (11.52) vimos que una fórmula para calcular la suma de cuadrado del error es:

$$SC_E = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} \tag{11.56}$$

Además, como la suma total de cuadrados, S_{yy}, está dada por:

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{(\Sigma_{i=1}^{n}\, y_i)^2}{n} = \mathbf{y}'\mathbf{y} - \frac{(\Sigma_{i=1}^{n}\, y_i)^2}{n}$$

La SC_E puede expresarse como:

$$\begin{aligned} SC_E &= \left[\mathbf{y}'\mathbf{y} - \frac{(\Sigma_{i=1}^{n}\, y_i)^2}{n}\right] - \left[\hat{\beta}'\mathbf{X}'\mathbf{y} - \frac{(\Sigma_{i=1}^{n}\, y_i)^2}{n}\right] \\ &= S_{yy} - SC_R \end{aligned}$$

Así, hemos obtenido una forma explícita para la suma de cuadrados de la regresión:

$$SC_R = \hat{\beta}'\mathbf{X}'\mathbf{y} - \frac{(\Sigma_{i=1}^{n}\, y_i)^2}{n} \tag{11.57}$$

El procedimiento de análisis de varianza para el modelo de regresión lineal múltiple se sintetiza en la tabla 11.11.

Coeficiente de determinación

El que un modelo sea significativo no necesariamente implica que sea bueno en términos de que explique la variación de los datos. Por ello es importante tener medi-

Tabla 11.11 ANOVA para la significancia del modelo de regresión lineal múltiple.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor-p
Regresión	$SC_R = \hat{\beta}'\mathbf{X}'\mathbf{y} - \frac{(\Sigma_{i=1}^{n} y_i)^2}{n}$	k	CM_R	CM_R/CM_E	$\Pr(F > F_0)$
Error o residuo	$SC_E = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y}$	$n-k-1$	CM_E		
Total	$S_{yy} = \mathbf{y}'\mathbf{y} - \frac{(\Sigma_{i=1}^{n} y_i)^2}{n}$	$n-1$			

ciones adicionales de la calidad del ajuste del modelo, como las gráficas de residuales y el coeficiente de determinación. Con la información del análisis de varianza de la tabla 11.11 es muy sencillo calcular el coeficiente de determinación, R^2, y el coeficiente de determinación ajustado, R^2_{aj}:

$$R^2 = \frac{SC_R}{S_{yy}} = 1 - \frac{SC_E}{S_{yy}}$$

$$R^2_{aj} = \frac{S_{yy}/(n-1) - CM_E}{S_{yy}/(n-1)} = \frac{CM_{\text{total}} - CM_E}{CM_{\text{total}}}$$

$$= 1 - \frac{CM_E}{CM_{\text{total}}}$$

Ambos coeficientes se interpretan de forma similar al caso de regresión lineal simple, es decir, como el porcentaje de variabilidad de los datos que son explicados por el modelo. Se cumple que $0 < R^2_{aj} \leq R^2 < 1$; en general, para hablar de un modelo que tiene un ajuste satisfactorio es necesario que ambos coeficientes tengan valores superiores a 0.7. Cuando en el modelo hay términos que no contribuyen de manera significativa a éste, el R^2_{aj} tiende a ser menor que el R^2. Por lo tanto, es deseable depurar el modelo y para ello las siguientes pruebas de hipótesis son de mucha utilidad.

Coeficiente de correlación múltiple
Es la raíz cuadrada del coeficiente de determinación R^2, y mide la intensidad de la relación entre la variable dependiente y las variables o términos en el modelo.

Coeficiente de correlación múltiple

Es la raíz cuadrada del coeficiente de determinación R^2:

$$R = \sqrt{R^2}$$

y es una medida de la intensidad de la relación entre la variable dependiente, Y, y el conjunto de variables o términos en el modelo ($X_1, X_2, \ldots, X_k$).

Error estándar de estimación y media del error absoluto

Al igual que en regresión lineal simple, el error estándar de estimación y la media del error absoluto proporcionan dos medidas del error de ajuste de un modelo, éstas tienen una interpretación similar a la que se dio para el caso de regresión lineal simple

(véase ecuaciones 11.36 y 11.37). En cuanto al cálculo en el caso múltiple, la $mea = (\Sigma_{i=1}^{n}|e_i|)/n$ y el error estándar de estimación, $\hat{\sigma} = \sqrt{SC_E/(n-k-1)}$.

Pruebas sobre coeficientes individuales del modelo

Un aspecto clave en un análisis de regresión múltiple es valorar qué tanto contribuye cada término a la explicación de la variable de respuesta, para de esa forma eliminar los que tienen una contribución poco importante o quizá pensar en agregar otras variables no consideradas. Las hipótesis para probar la significancia de cualquier coeficiente individual, β_j, se especifica de la siguiente manera:

$$\begin{aligned} H_0 &: \beta_j = 0 \\ H_A &: \beta_j \neq 0 \qquad j = 0, 1, 2, \ldots, k \end{aligned} \tag{11.58}$$

De acuerdo con la sección anterior, el estimador de mínimos cuadrados $\hat{\beta}$ es un vector aleatorio, cuya distribución es normal con media β y matriz de covarianza $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. De aquí que la distribución de los coeficientes de regresión $\hat{\beta}_j$ sea:

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 C_{j+1,j+1})$$

donde $C_{j+1,j+1}$ es el elemento de la diagonal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ correspondiente al parámetro $\hat{\beta}_j$. De aquí, y dado que σ^2 se estimó con el CM_E (ecuación 11.53), entonces el estadístico de prueba para examinar la hipótesis de la expresión (11.58) está dado por

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{CM_E C_{j+1,j+1}}} \tag{11.59}$$

donde se rechaza H_0 si $|t_0| > t_{(\alpha/2,\, n-k-1)}$ o en forma equivalente si valor-$p = P(T > |t_0|) < \alpha$. En la tabla 11.12 se muestra un resumen del análisis sobre el modelo de regresión basado en la prueba antes descrita.

Ejemplo 11.4

Aplicamos las pruebas y cálculos descritos en esta sección a los datos del ejemplo 11.3; en la tabla 11.13 se muestran los resultados obtenidos, para el análisis del modelo de regresión. Recordemos que se ajustó un modelo de segundo orden: $Y = -75\,732.8 + 4\,438.69X_1 + 5\,957.79X_2 - 17.9688X_1X_2 - 181.316\,X_1^2 - 146.404\,X_2^2$. A partir de esta tabla vemos que el único término que no es significativo, de acuerdo con la prueba t, es la interacción X_1X_2 y los términos que tienen una mayor contribución a la respuesta son X_2 y X_2^2. En el análisis de varianza se aprecia que el modelo de regresión es significativo, y de acuerdo con los coeficientes de determinación, R^2 y R_{aj}^2, el modelo explica bien la variabilidad presente en los datos. También se aprecia el error estándar de estimación, $\hat{\sigma}$, y la media del error absoluto, que dada la escala de medición de la variable de respuesta, éstos tienen una magnitud relativamente pequeña.

Tabla 11.12 Análisis de regresión para el modelo $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$.

Parámetro	Estimación	Error estándar	Estadístico	Valor-*p*
Intercepción	$\hat{\beta}_0$	$\sqrt{CM_E C_{11}}$	$\frac{\hat{\beta}_0}{\sqrt{CM_E C_{11}}}$	$\Pr(T > \lvert t_0 \rvert)$
β_1	$\hat{\beta}_1$	$\sqrt{CM_E C_{22}}$	$\frac{\hat{\beta}_1}{\sqrt{CM_E C_{22}}}$	$\Pr(T > \lvert t_0 \rvert)$
⋮	⋮	⋮	⋮	⋮
β_k	$\hat{\beta}_k$	$\sqrt{CM_E C_{k+1,k+1}}$	$\frac{\hat{\beta}_k}{\sqrt{CM_E C_{k+1,k+1}}}$	$\Pr(T > \lvert t_0 \rvert)$

Selección de variables en regresión lineal múltiple

El procedimiento descrito para probar la significancia de un coeficiente individual en el modelo no es completamente decisiva, debido a que en ocasiones los estimadores $\hat{\beta}_j$ no son independientes entre sí, como se puede apreciar en los elementos fuera de la diagonal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Esto hace que un coeficiente β_h aparente ser significativo porque su estimador está correlacionado con el estimador, $\hat{\beta}_j$, de otro coeficiente que sí tiene una contribución significativa.

A pesar de lo anterior, la prueba *t* sobre la significancia de los términos del modelo, combinada con los coeficientes de determinación y el error cuadrático medio, $\sqrt{CM_E}$, puede ser de utilidad para tener un modelo depurado en el cual la mayoría de los términos realmente ayuden a explicar la variable de respuesta. Un procedimiento con apoyo de un software estadístico sería el siguiente:

Tabla 11.13 Análisis para el modelo de regresión lineal múltiple ajustado a los datos del ejemplo 11.3.

Parámetro	Estimación	Error estándar	Estadístico	Valor-*p*
Constante	–75 732.8	6 313.95	–11.9945	0.0000
X_1: Carbono	4 438.69	708.101	6.2684	0.0008
X_2: Temperatura	5 957.79	347.095	17.1647	0.0000
X_1X_2	–17.9688	17.3848	–1.03359	0.3412
X_1^2	–181.316	27.488	–6.5962	0.0006
X_2^2	–146.404	6.87186	–21.3049	0.0000

Análisis de varianza

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor-*p*
Modelo	$5.51626 \times (10)^7$	5	$1.10325 \times (10)^7$	142.59	0.0000
Residual	464 228.0	6	77 371.3		
Total (Corr.)	$5.56268 \times (10)^7$	11			

$R^2 = 0.992$
$R^2_{aj} = 0.985$
Error estándar de estimación = 278.157
Media del error absoluto = 145.324

- En una tabla en la que se muestre la prueba t para todos los términos se elige el que tuvo menor contribución (esto se aprecia en el valor más pequeño en términos absolutos de su correspondiente estadístico de prueba, t_0). Este término se quita del modelo, se ajusta un nuevo modelo y se comparan los cambios en R^2, R^2_{aj} y $\sqrt{CM_E}$ para los dos modelos. Si los cambios son menores, es decir, si $\sqrt{CM_E}$ y R^2 prácticamente quedan igual, y quizás el R^2_{aj} sube un poco, entonces ese primer término se puede eliminar definitivamente del modelo.
- Al modelo ajustado sin el término eliminado en el primer paso, se le aplica el paso anterior. Este proceso continúa hasta que sólo queden en el modelo términos significativos.

La aplicación del procedimiento anterior ayuda a depurar el modelo aunque puede dejar términos que en realidad no contribuyan en forma importante al ajuste del modelo. Además, es viable si se tiene un modelo con pocos términos; pero, si se tiene un modelo con muchos términos es necesario recurrir a un procedimiento más robusto que proporcione mayor garantía para construir modelos que sólo tengan términos significativos.

Estas técnicas se conocen como *técnicas de selección de variables por pasos*. Por ejemplo, la técnica *hacia delante* (*forward*), en el primer paso compara todos los posibles modelos de una sola variable, y se queda con el modelo que logra una mejor explicación. La variable o término que corresponde a este primer mejor modelo se convierte en parte del modelo final. En el segundo paso se inicia con el modelo del paso anterior y se le agrega el segundo término que mejor ayude a mejorar el ajuste del modelo. Este proceso continúa hasta que llega el momento en que agregar otro término no mejora de manera significativa el modelo. Otra técnica de selección de variables se conoce como *hacia atrás* (*backward*), y tiene un proceso inverso al descrito antes: inicia con un modelo que incluye todos los términos, y paso a paso le va quitando las variables que menos contribuyen al ajuste. La forma de agregar o quitar variables se basa en pruebas F y no en la prueba t. Obviamente estas técnicas de selección de variables tienen una alta demanda de cálculo. La mayoría de las aplicaciones de software estadístico incluye estas técnicas de selección de variables.

En resumen, si se cuenta con un modelo con muchos términos y se quiere tener un modelo depurado que incluya sólo términos que realmente ayuden a explicar la variable de respuesta, entonces se recomienda aplicar una técnica de selección de variables como la descrita antes, apoyándose desde luego en un software estadístico. Pero, si se tiene un modelo con pocos términos, entonces es posible aplicar una técnica de selección de variables, o bien, recurrir al procedimiento basado en la prueba t sobre los coeficientes individuales que se describió antes.

Intervalos de confianza y predicción en regresión múltiple

Al igual que en regresión lineal simple es posible construir intervalos de confianza y predicción en regresión lineal múltiple. Por ejemplo, a partir de la tabla 11.12 es

claro que un estimador por intervalo de cada coeficiente en lo individual está dado por:

$$\hat{\beta}_j - t_{(\alpha/2,\, n-k-1)}\sqrt{CM_E C_{j+1,\, j+1}} < \beta_j < \hat{\beta}_j + t_{(\alpha/2,\, n-k-1)}\sqrt{CM_E C_{j+1,\, j+1}}$$

También es posible obtener un intervalo de confianza con respecto a la respuesta media en un punto particular, digamos $x_{10}, x_{20}, \ldots x_{k0}$. Si definimos el vector:

$$\mathbf{x}_0^1 = \begin{bmatrix} 1 & x_{10} & x_{20} & \cdots & x_{k0} \end{bmatrix}$$

entonces la respuesta media estimada en este punto es:

$$\hat{\mathbf{y}}_0 = \mathbf{x}_0'\hat{\beta}$$

Se trata de un estimador insesgado cuya varianza es:

$$V(\hat{\mathbf{y}}_0) = \sigma^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$$

Por lo tanto, un intervalo de confianza del $100(1 - \alpha)\%$ de confianza para la respuesta media en el punto $x_{10}, x_{20}, \ldots x_{k0}$ está dado por:

$$\hat{\mathbf{y}}_0 - t_{(\alpha/2,\, n-k-1)}\sqrt{CM_E \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} < E(\mathbf{y}_0) < \hat{\mathbf{y}}_0 + t_{(\alpha/2,\, n-k-1)}\sqrt{CM_E \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \quad \textbf{(11.60)}$$

Una de las aplicaciones más frecuentes del análisis de regresión es predecir observaciones futuras con base en el modelo de regresión lineal múltiple. Si deseamos predecir un valor futuro de la variable de respuesta, $\mathbf{y}_0$, en un nivel particular de las variables regresoras, digamos $x_{10}, x_{20}, \ldots, x_{k0}$. Si $\mathbf{x}_0' = (1, x_{10}, x_{20}, \ldots, x_{k0})$, entonces una estimación puntual de la observación futura $\mathbf{y}_0$ en el nivel referido está dada por:

$$\hat{\mathbf{y}}_0 = \mathbf{x}_0'\hat{\beta}$$

En forma similar al caso de regresión lineal simple se puede demostrar que el intervalo de predicción al $100(1 - \alpha)\%$ para $\mathbf{y}_0$ es:

$$\hat{\mathbf{y}}_0 - t_{(\alpha/2,\, n-k-1)}\sqrt{CM_E\left(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right)} < \mathbf{y}_0 < \hat{\mathbf{y}}_0 + t_{(\alpha/2,\, n-k-1)}\sqrt{CM_E\left(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right)} \quad \textbf{(11.61)}$$

Como ya se dijo, al estimar por intervalo la respuesta media con la ecuación 11.60 o predecir una nueva observación con la ecuación 11.60 se debe tener cuidado de hacer extrapolaciones fuera de la región de las observaciones con las que se ajustó el modelo. Fuera de la región, los aspectos físicos o sociales que están atrás de todo modelo de regresión pueden empezar a actuar de otra forma.

Uso de un software estadístico

La mayoría de los sistemas computacionales especializados en estadística incluyen procedimientos para realizar análisis de regresión tanto simple como múltiple, y por lo general incluyen técnicas de selección de variables. Por ejemplo, en *Statgraphics* todas estas herramientas se incluyen en *Relate*, y dentro se encuentran los siguientes procedimientos:

- *Simple Regression Analysis*
- *Polynomial Regression Analysis*
- *Box-Cox Transformation Analysis*
- *Multiple Regression Analysis*

Por ejemplo en la opción *polynominal* se pueden ajustar modelos del tipo $y = \beta_0 + \beta_1 x + \beta_2 x^2$. Al accesar a cualquiera de esos procedimientos, la interacción con el software es bastante clara. De la versión *Centurion* en adelante se agregan otros modelos de regresión, como por ejemplo para datos de atributos. En *Minitab* se accede a varios procedimientos de regresión con la secuencia: *Stat* → *Regression*.

Excel

En la hoja de cálculo de Excel se incluye la regresión lineal simple y múltiple; para ello, es necesario realizar la siguiente secuencia de opciones: *Herramientas* → *Complementos*, y asegurarse de que ahí esté activada la opción correspondiente a *Herramientas para análisis*. Después se debe seguir secuencia de procedimientos:

Herramientas → *Análisis de datos* → *Regresión*

Enseguida se solicitará el rango de celdas donde se encuentran los datos para la variable dependiente —*Rango Y de entrada*— y para la(s) variable(s) regresora(s) —*Rango X de entrada*—. En caso de que se tenga más de una variable, o incluso un modelo con interacciones o términos cuadráticos, entonces hay que darle todo el rango donde se encuentran los datos correspondientes a la matriz *X* (véase ejemplo 11.3). Después habrá que activar las casillas según las características del análisis que se deseen. Por ejemplo, el *Nivel de confianza deseado*; mientras que *Constante igual a cero* es una casilla que se activa para que la línea de regresión pase por el origen, es decir, si se quiere que el modelo no incluya el parámetro β_0.

Preguntas y ejercicios

1. ¿Cuál es el propósito general del análisis de regresión?
2. En el análisis de regresión intervienen dos tipos de variables: las independientes y las dependientes. Explique con sus palabras y a través de ejemplos, las características de estos dos tipos de variables.

3. Considere el modelo de regresión lineal simple, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$; con $i = 1, 2, \ldots, n$, y suponiendo que para estimar los parámetros se utilizaron un total de 10 observaciones, es decir, $n = 10$, conteste las siguientes preguntas:
   a) Suponga una buena relación lineal entre las variables *X* y *Y*; construya un diagrama de dispersión hipotético que refleje esta relación.
   b) Sobre el diagrama de dispersión anterior, ajuste a "ojo" la mejor línea recta que describa la relación observada.
   c) Utilice el procedimiento de mínimos cuadrados y explique en forma esquemática el procedimiento matemático para estimar los parámetros que minimizan los errores.
   d) Explique el significado de los dos parámetros del modelo (β_0 y β_1).
   e) Escriba las expresiones que estiman a los dos parámetros del modelo.
   f) ¿Cuáles son las suposiciones que se hacen sobre los errores (ε_i)?

4. Considere el modelo de regresión lineal simple, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, conteste:
   a) Formule las hipótesis que se hacen sobre los parámetros del modelo y explique la consecuencia de aceptar o rechazar cada una de éstas.
   b) Anote en forma detallada el estadístico de prueba, t_0, para cada una de las hipótesis y dé una explicación de por qué sirven para probar las hipótesis. Es decir, determine cuándo estos estadísticos tienen valores pequeños o grandes, y la decisión que se tomaría con respecto a su hipótesis correspondiente.
   c) Con respecto al análisis de varianza para el modelo, escriba y explique la hipótesis correspondiente. Además, anote con detalle el estadístico de prueba, F_0, y dé una justificación de por qué tal estadístico sirve para probar tal hipótesis.

5. Con respecto a los intervalos de confianza para la recta y los intervalos de predicción, señale cómo se obtienen y para qué se aplica cada uno de ellos.

6. En una etapa inicial del procesamiento mecánico de piezas de acero, se sabe que una herramienta sufre un deterioro gradual que se refleja en cierto diámetro de las piezas manufacturadas. Para predecir el tiempo de vida útil de la herramienta se tomaron datos de horas de uso y el diámetro promedio de cinco piezas producidas al final de la jornada. Los datos obtenidos para una herramienta se muestran a continuación:

Horas de uso	Diámetro (mm)
16	26.2
32	25.7
48	26.0
64	27.7
80	28.3
96	29.5
112	30.1
128	31.8
144	31.4
160	33.4
176	33.6
192	32.7
208	35.0
224	36.1
240	35.7
256	36.2
272	36.8
288	39.1
304	38.7
320	39.2

a) ¿En este problema cuál variable se considera independiente y cuál dependiente?
b) Mediante un diagrama de dispersión analice la relación entre estas dos variables. ¿Qué tipo de relación observa y cuáles son algunos hechos especiales?
c) Haga un análisis de regresión (ajuste una línea recta a estos datos, aplique pruebas de hipótesis y verifique residuos).
d) ¿La calidad del ajuste es satisfactoria? Argumente.
e) Si el diámetro máximo tolerado es de 45, ¿cuántas horas de uso estima que tiene esa herramienta?
f) Señale el valor de la pendiente de la recta e interprételo en términos prácticos.
g) Obtenga el error estándar de estimación y comente qué relación tiene éste con la calidad del ajuste.

7. En un proceso de extracción se estudia la relación entre tiempo de extracción y rendimiento. Los datos obtenidos se muestran en la siguiente tabla.

Tiempo (minutos)	Rendimiento (%)
10	64
15	81.7
20	76.2
8	68.5
12	66.6
13	77.9
15	82.2
12	74.2
14	70
20	76
19	83.2
18	85.3

a) ¿En este problema cuál variable se considera independiente y cuál dependiente?
b) Mediante un diagrama de dispersión analice la relación entre estas dos variables. ¿Qué tipo de relación observa y cuáles son algunos hechos especiales?
c) Haga un análisis de regresión (ajuste una línea recta a estos datos, aplique pruebas de hipótesis y verifique residuos).
d) ¿La calidad del ajuste es satisfactoria? Argumente.
e) Destaque el valor de la pendiente de la recta e interprételo en términos prácticos.
f) Estime el rendimiento promedio que se espera a un tiempo de extracción de 25 minutos y obtenga un intervalo de confianza para esta estimación.

8. En cierta empresa es usual pagar horas extra para cumplir con los tiempos de entrega. En este centro productivo un grupo de mejora de calidad trata de reducir la proporción de piezas malas, para ello deciden investigar la relación entre la cantidad de horas extra, X, y el porcentaje de artículos defectuosos, Y. En la siguiente tabla se muestran los datos obtenidos.

Semana	Horas extra	Defectos %
1	340	5
2	95	3
3	210	6
4	809	15
5	80	4
6	438	10
7	107	4
8	180	6
9	100	3
10	550	13
11	220	7
12	50	3
13	193	6
14	290	8
15	340	2
16	115	4
17	362	10
18	300	9
19	75	2
20	93	2
21	320	10
22	154	7

a) ¿De estas variables cuál se puede suponer independiente y cuál dependiente?
b) Obtenga el diagrama de dispersión para estas variables. ¿Qué relación observa?
c) Haga un análisis de regresión (ajuste una línea recta a estos datos, aplique pruebas de hipótesis y verifique residuos).
d) ¿La calidad del ajuste es satisfactoria?
e) El que dos variables estén relacionadas no necesariamente implica que haya una relación causa-efecto. Sin embargo, a pesar de esto, ¿puede concluir con seguridad que cuando se trabaja tiempo extra se incrementa el porcentaje de defectuosos, porque ocurren factores como calentamiento de equipo, cansancio de obreros, etc., y todo esto causa mayores problemas en la calidad de las piezas?

9. En una industria se desea investigar cómo influye la temperatura (°C) en la presión del vapor de B-trimetilboro, los datos obtenidos para tal propósito se muestran en la siguiente tabla.

Temperatura	Presión
13	2.9
19.5	5.1
45.7	30.5
56.1	51.4
64.4	74.5
71.4	100.2
80.5	143.7
85.7	176.9
22.5	8.5
27.2	10.3
31.8	14.6

a) Construya un diagrama de dispersión e interprételo.
b) Ajuste una línea recta y observe la calidad de ajuste.
c) Señale el valor de la pendiente de la recta e interprételo en términos prácticos.
d) Observe la gráfica de residuales contra predichos, ¿nota algo relevante?
e) ¿Está satisfecho con el modelo ajustado? Argumente.
f) ¿Hay algún otro modelo que puede funcionar mejor? Proponga uno de los que se explicaron al inicio del capítulo.

10. En un proceso de manufactura se utiliza una herramienta de corte y se quiere investigar la relación entre la velocidad de corte (metros por minuto) y el tiempo de vida (horas) de la herramienta. Los datos obtenidos para esta investigación se muestran a continuación:

Velocidad	Vida
20	8.7
20	9.5
25	8.5
25	7.7
25	8.4
30	8.0
30	5.3
30	7.3
35	7.8
35	5.7
35	6.1
40	4.3
40	4.2

a) Mediante un diagrama de dispersión analice la relación entre estas dos variables. ¿Qué tipo de relación observa?
b) Haga un análisis de regresión (ajuste una línea recta a estos datos, aplique pruebas de hipótesis y verifique residuos).
c) ¿La calidad del ajuste es satisfactoria? Argumente.
d) Si normalmente la herramienta se opera a una velocidad de 30 metros por minuto, estime el tiempo medio de vida tanto de manera puntual como por intervalo.
e) Señale el valor de la pendiente de la recta e interprételo en términos prácticos.
f) Obtenga el error estándar de estimación y comente qué relación tiene con la calidad del ajuste.

11. A partir de la siguiente tabla de datos realice los cálculos necesarios y complete una tabla similar a la 11.2.

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
0	4			
1	3			
2	6			
3	9			
4	9			
5	11			
6	12			
7	14			
$\sum_{i=1}^{n} x_i =$	$\sum_{i=1}^{n} y_i =$	$\sum_{i=1}^{n} x_i y_i =$	$\sum_{i=1}^{n} x_i^2 =$	$\sum_{i=1}^{n} y_i^2 =$

a) Realice los cálculos indicados en la tabla.
b) Con base en lo anterior, construya la tabla de análisis de regresión para la recta de regresión (tabla 11.4) y el análisis de varianza (tabla 11.5).
c) A partir de lo anterior obtenga conclusiones.
d) Obtenga el coeficiente de determinación y valore la calidad del ajuste.

12. Como parte del análisis del problema de ausentismo, se decide investigar la relación entre edad del empleado y días que faltó a laborar en el año. Los datos del último año se muestran en la siguiente tabla:

Empleado	Edad	Faltas	Empleado	Edad	Faltas
1	29	6	21	25	7
2	33	5	22	38	3
3	40	0	23	22	0
4	23	8	24	30	4
5	31	6	25	24	7
6	20	9	26	39	10
7	30	5	27	35	5
8	38	6	28	20	1
9	23	8	29	32	5
10	25	6	30	25	5
11	26	7	31	36	5
12	30	5	32	30	5
13	42	2	33	20	10
14	34	5	34	38	4
15	31	6	35	39	4
16	18	11	36	34	4
17	33	6	37	35	6
18	33	4	38	27	7
19	33	5	39	40	3
20	32	5	40	31	6

a) ¿En este problema cuál variable se puede ver como independiente y cuál como dependiente?
b) Mediante un diagrama de dispersión analice la relación entre estas dos variables.
c) ¿Qué tipo de relación observa y cuáles son algunos hechos especiales?
d) Haga un análisis de regresión (ajuste una línea recta a estos datos, aplique pruebas de hipótesis y verifique residuales).
e) ¿La calidad del ajuste es satisfactoria? Argumente.
f) ¿Cuál es el error estándar de estimación?

13. Elimine los cuatro datos que le parezcan más atípicos en el problema anterior, y repita los incisos *d*), *e*) y *f*). ¿Los resultados obtenidos son diferentes?

14. En una fábrica de pintura se quiere reducir el tiempo de secado del barniz. Los siguientes datos corresponden al tiempo de secado del barniz (en horas) y a la cantidad de aditivo con el que se intenta lograr tal reducción.

Cantidad de aditivo	Tiempo de secado
0	14
1	11
2	10
3	8
4	7.5
5	9
6	10
7	11
8	13
9	12
10	15

a) Mediante un diagrama de dispersión investigue la relación entre el tiempo de secado y la cantidad de aditivo.
b) Con base en la relación, ¿qué cantidad de aditivo recomendaría para reducir el tiempo de secado?
c) Obtenga el coeficiente de correlación entre ambas variables e interprételo.
d) Al parecer, el coeficiente de correlación lineal es muy bajo, ¿esto significa que el tiempo de secado no está relacionado con la cantidad de aditivo?
e) ¿Usted cree que sea correcto ajustar una línea recta?
f) Proponga el modelo que crea adecuado, ajústelo y haga un análisis de regresión completo para tal modelo.

Preguntas y ejercicios de regresión lineal múltiple

15. ¿Por qué se requiere la regresión lineal múltiple?

16. Considere el siguiente modelo $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \varepsilon$, ¿es un modelo de regresión lineal múltiple? Argumente.

17. Considere un modelo de regresión lineal múltiple con cuatro variables: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{4i} + \varepsilon_i$; $i = 1, 2, \ldots, n$, y suponga que para estimar los parámetros se utilizaron un total de 12 observaciones, es decir, $n = 12$. Conteste las siguientes preguntas:
a) Explique en forma esquemática el procedimiento matemático para estimar los parámetros que minimizan los errores por mínimos cuadrados.
b) Denote el modelo en forma matricial: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, exprese con precisión todas las matrices involucradas en el modelo.
c) Proporcione la expresión matricial para los estimadores de mínimos cuadrados.
d) Especifique la hipótesis de significancia del modelo y lo que significa aceptar o rechazar esta hipótesis.
e) Dé la expresión del estadístico de prueba, F_0, para la hipótesis anterior, así como una explicación racional de por qué funciona como estadístico de prueba, es decir, vea cuándo este estadístico tiene valores grandes o pequeños, y lo que eso significa en términos de calidad de ajuste.
f) Formule las hipótesis sobre los parámetros individuales del modelo y comente qué significa aceptar o rechazar cada una de éstas.
g) Proporcione la expresión para el estadístico de prueba para el caso anterior y comente por qué estos estadísticos funcionan como criterio de aceptación o rechazo.

h) ¿Cuáles son los riesgos de hacer predicciones fuera de la región de los datos originales?

18. En una empresa dedicada a anodizar artículos de aluminio (baterías de cocina), el anodizado se logra con una solución hecha a base de ácidos (sulfúrico, cítrico, bórico) y dicromato de aluminio. En este proceso se controla el pH de la solución, la temperatura, la corriente y el tiempo de permanencia. Debido al poco grosor del anodizado, han aumentado las quejas por la escasa resistencia y durabilidad del producto. Para resolver este problema se decide estudiar, mediante un experimento, la relación del pH y la temperatura con el grosor del anodizado. Los datos se muestran en la siguiente tabla:

pH	Temperatura	Espesor
1.2	–8	9
1.8	–8	14
1.2	8	10
1.8	8	19
1.2	–8	8
1.8	–8	12
1.2	8	11
1.8	8	20
1.5	0	14
1.5	0	13

a) ¿Cuáles son las variables independientes y cuál la dependiente? Argumente.
b) Ajuste un modelo del tipo $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ y anote la ecuación del modelo ajustado.
c) A partir del modelo ajustado, ¿cuál es el espesor estimado cuando se utiliza un pH = 2 y una temperatura de 10 grados?
d) ¿El modelo es adecuado? Argumente con base en gráficas de residuos, pruebas de hipótesis y coeficientes de determinación.
e) ¿Cree que valdría la pena pensar en añadir otro término al modelo para mejorar el ajuste? Argumente.

19. Repita el problema anterior, pero ahora ajustando el modelo $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon$.

20. Ajuste a los datos del ejemplo 11.3 un modelo que no incluya el término $X_1 X_2$; es decir, ajuste el modelo $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \varepsilon$. Haga un análisis completo y compare la calidad del ajuste con lo que se realizó en el ejemplo 11.4.

21. Considere los datos que aparecen en la siguiente tabla:

y	x_1	x_2
6	10	3
9	12	11
8	12	4
3	4	1
10	12	11
4	6	1
5	8	7
2	2	4
11	18	8
9	9	10
10	17	8
2	2	5

a) Ajuste el siguiente modelo $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, es decir, encuentre los estimadores de mínimos cuadrados para estos coeficientes de regresión.
b) A partir del modelo ajustado, estime la respuesta media cuando $x_1 = 8$ y $x_2 = 7$; ¿este valor es diferente al observado en las mismas condiciones? De ser así, ¿por qué ocurre esto?
c) Haga la estimación por intervalo para la respuesta media en el punto anterior.
d) Construya un intervalo de predicción para una observación futura teniendo $x_1 = 8$ y $x_2 = 7$.
e) Explique las diferencias entre los dos intervalos anteriores.
f) ¿Las estimaciones anteriores son adecuadas? Argumente con base en la calidad de ajuste del modelo.

22. Se realizó un experimento para estudiar el sabor del queso panela en función de la cantidad del cuajo y la sal. La variable de respuesta observada es el sabor promedio reportado por un grupo de cinco panelistas que probaron todos los quesos y los calificaron con una escala hedónica. Los datos obtenidos se muestran a continuación:

Sal	Cuajo	Sabor
6	0.3	5.67
5.5	0.387	7.44
4.5	0.387	7.33
4	0.3	6.33
4.5	0.213	7.11
5.5	0.213	7.22
5	0.3	6.33
5	0.3	6.66

a) Ajuste el modelo $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$.
b) ¿El modelo explica la variación observada en el sabor? Argumente con base en la significancia del modelo, los residuales y el coeficiente de determinación.
c) Ajuste un modelo que incluya términos cuadráticos y analice con detalle la calidad del ajuste.
d) Compare el error estándar de estimación $\left(\sqrt{CM_E}\right)$ y los coeficientes de determinación (R^2 y R^2_{aj}) para ambos modelos.
e) ¿Cuál modelo prefiere para explicar el sabor?

23. En el área de desarrollo de una empresa se pretende obtener un nuevo polímero de bajo peso molecular (Y_1), de lograrse esto, se obtendrá un polímero que funcione como dispersante en la industria de la cerámica. De acuerdo con los conocimientos técnicos que se tienen, se considera que los factores críticos son X_1: persulfato de sodio (NaPS), X_2: ácido hipofosforoso (H_3PO_2) y X_3: isopropanol (IPA). Para encontrar las condiciones óptimas se realizó un experimento y se obtuvieron los siguientes datos (los valores de los factores están codificados). Además de Y_1 se midió la viscosidad (Y_2).

X_1	X_2	X_3	Y_1	Y_2
0	0	0	8 392	1 075
–1	–1	0	9 895	2 325
1	–1	0	9 204	1 575
–1	1	0	7 882	690
1	1	0	7 105	420
–1	0	–1	8 939	1 188
1	0	–1	8 548	930
0	0	0	8 598	920
–1	0	1	9 152	1 275
1	0	1	8 992	860
0	–1	–1	10 504	5 600
0	1	–1	7 462	540
0	–1	1	9 368	1 225
0	1	1	7 772	620
0	0	0	8 440	1 015

a) Ajuste el modelo $Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ para la variable Y_1.

b) ¿El modelo explica la variación observada en Y_1? Argumente con base en la significancia del modelo, los residuales y los coeficientes de determinación.

c) Ajuste el modelo $Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + \varepsilon$, y analice con detalle la calidad del ajuste (hipótesis sobre coeficientes individuales, gráficas de residuos).

d) Compare el error estándar de estimación $\left(\sqrt{CM_E}\right)$ y los coeficientes de determinación (R^2 y R^2_{aj}) para ambos modelos.

e) Con base en lo anterior, proponga un modelo que considere que sólo tiene términos significativos. Ajústelo y haga un análisis completo sobre éste.

f) Para el modelo final al que llegó en el punto anterior, interprete con detalle el significado de cada uno de los coeficientes estimados en función de su aporte para la variable de respuesta Y_1.

24. Realice el ejercicio anterior pero ahora para la otra variable, Y_2. Destaque similitudes y diferencias.

25. Se tiene un proceso de extrusión para producir harina instantánea de amaranto. Una de las variables que interesa minimizar es el *índice de solubilidad en agua* (ISA) y los factores que se controlan son: *temperatura* (X_1), *porcentaje de humedad* (X_2) y *velocidad de tornillo* (X_3). Con las variables independientes codificadas, los datos obtenidos mediante un diseño de experimentos Box-Behnken se muestran a continuación:

x_1	x_2	x_3	*ISA*
–1	–1	0	15.87
1	–1	0	12.70
–1	1	0	14.80
1	1	0	13.53
–1	0	–1	15.10
1	0	–1	12.47
–1	0	1	11.37
1	0	1	10.27
0	–1	–1	15.33
0	1	–1	15.53
0	–1	1	15.17
0	1	1	14.17
0	0	0	13.85
0	0	0	13.93
0	0	0	13.77

a) Ajuste un modelo de regresión lineal múltiple que incluya las tres variables independientes, interacciones y términos cuadráticos.
b) Haga un análisis detallado sobre el modelo ajustado (hipótesis sobre los coeficientes individuales, gráficas de residuos y coeficientes de determinación).
c) Estime la respuesta en $x_1 = 1, x_2 = 0, x_3 = -1$.
d) Haga una estimación por intervalo para la respuesta media en el punto anterior y también una predicción por intervalo para una nueva observación.
e) ¿El modelo es adecuado? Si usted cree que un modelo más simple podría lograr resultados similares en cuanto a la calidad de ajuste, haga las exploraciones necesarias y proponga un modelo final.
f) Compare su modelo final con el modelo más general que ajustó al principio en términos de los coeficientes de determinación y del error estándar de estimación. ¿Qué aprecia?

Capítulo 12

Optimización de procesos con metodología de superficie de respuesta

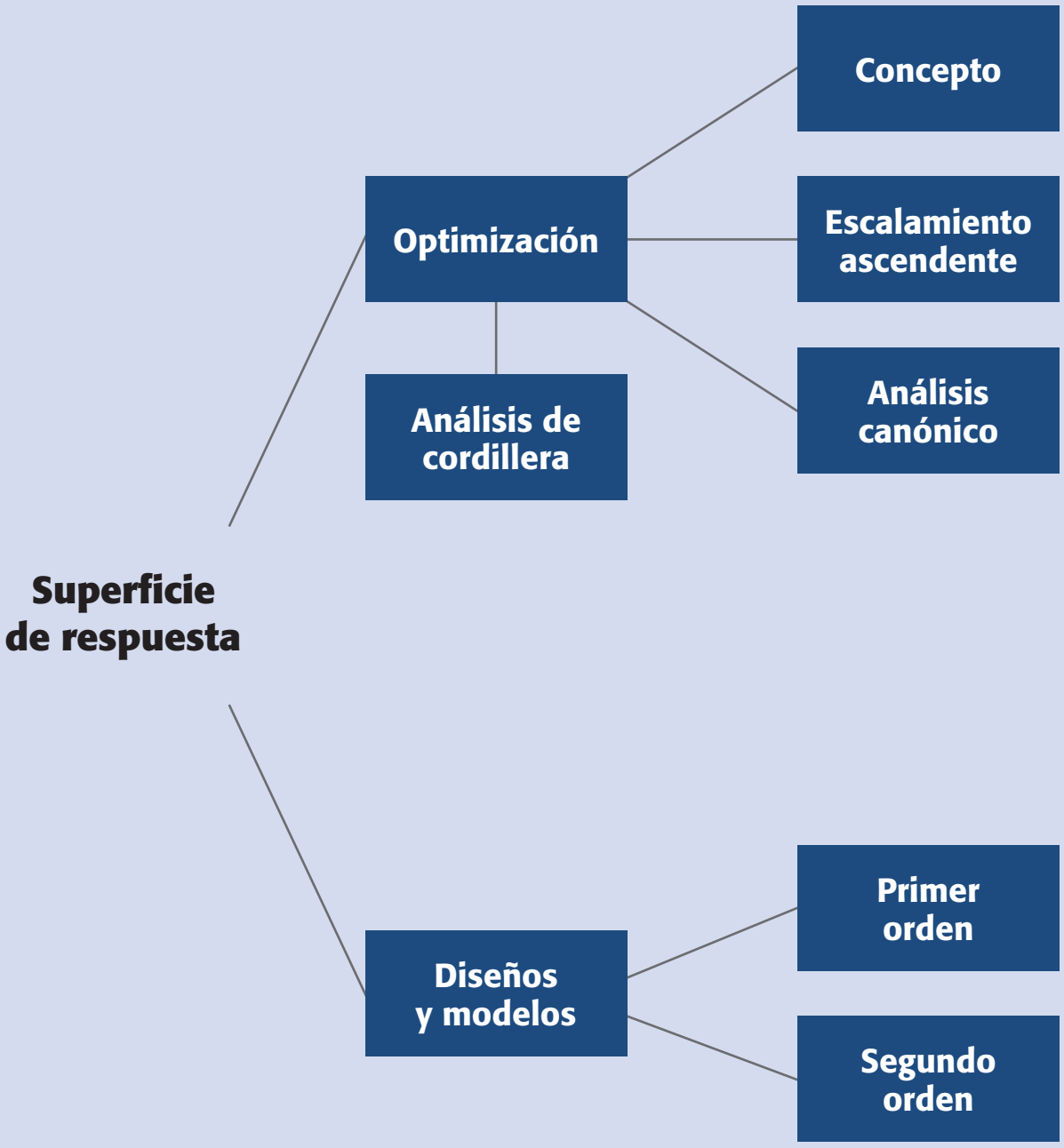
Sumario

- Introducción a la metodología de superficie de respuesta
- Técnicas de optimización
- Diseños de superficie de respuesta
- Uso de software estadístico

Objetivos de aprendizaje

- Explicar el concepto de optimización y su relación con la superficie de respuesta.
- Aplicar la metodología de superficie de respuesta y sus respectivos diseños y modelos.
- Describir las técnicas de optimización y aplicarlas adecuadamente.

Mapa conceptual



Conceptos clave

- Análisis canónico
- Análisis de cordillera
- Búsqueda de primer orden
- Búsqueda de segundo orden
- Cordillera estacionaria
- Cresta ascendente
- Cribado
- Diseño de composición central
- Diseño de primer orden
- Diseño de segundo orden
- Diseño en MSR
- Diseño ortogonal
- Diseño rotable
- Diseño simplex
- Diseños de Box-Behnken
- Diseños de superficie de respuesta
- Escalamiento ascendente
- Mejor tratamiento
- Metodología de la superficie de respuesta
- Modelo en MSR
- Operación evolutiva
- Optimización
- Punto estacionario
- Punto óptimo
- Región de operabilidad
- Región experimental

Introducción a la metodología de superficie de respuesta

Como se explicó en el capítulo anterior, algunas veces hay experimentos con los que no se obtienen las respuestas buscadas o el nivel de mejoras logrado no es suficiente, por lo que es necesario experimentar de manera secuencial hasta encontrar el nivel de mejoras deseado. En este caso, después de una primera etapa experimental quizá sea necesario desplazar la región experimental (moverse de lugar) en una dirección adecuada, o bien, explorar en forma más detallada la región experimental inicial (véase figura 12.1). La forma de realizar ambas cosas son parte de la llamada *metodología de superficie de respuesta*[1] (MSR).

La MSR es la estrategia experimental y de análisis que permite resolver el problema de encontrar las condiciones de operación óptimas de un proceso, es decir, aquellas que dan por resultado "valores óptimos" de una o varias características de calidad del producto.

Región experimental y región de operabilidad

La *región experimental* es el espacio delimitado por los rangos de experimentación utilizados con cada factor. La *región de operabilidad* está delimitada por el conjunto de puntos o condiciones donde el equipo o proceso puede ser operado. Es difícil delimitar con certeza el tamaño de la región de operabilidad, ya que aun cuando se conozca (por especificaciones del equipo) el rango en que se puede colocar cada factor individual, es necesario determinar esos límites considerando varios factores de manera simultánea. Por ejemplo, es posible que la *temperatura* se pueda correr en su nivel más alto de operabilidad, siempre y cuando los factores *velocidad* y *fuerza* se mantengan bajos. La región de operabilidad considera todas las combinaciones posibles de los niveles de los factores donde el proceso puede operarse y ésta siempre es igual o más grande que la región experimental. Para mayor sencillez se consideran regiones de forma regular; en la figura 12.2 se representan las regiones de operabilidad y experimental.

En la MSR es importante tener presente esta visión de las regiones de operabilidad y experimental, ya que en principio, el punto óptimo que interesa encontrar pudiera localizarse en cualquier lugar de la región de operabilidad, dentro o afuera de la región experimental inicial. En procesos ya establecidos y muy estudiados, es de esperarse que dicho punto óptimo se encuentre "no muy lejos" de las condiciones de operación usuales, posiblemente dentro de la región experimental inicial. En cambio, cuando el proceso es nuevo o cuando se está escalando o rediseñando, es más probable que el punto de interés se ubique fuera de la primera región experimental

Metodología de la superficie de respuesta
Estrategia experimental y de modelación que permite encontrar condiciones de operación óptima de un proceso.

Región experimental
Espacio delimitado por los rangos de experimentación utilizados con cada factor.

Región de operabilidad
Conjunto de condiciones donde el equipo o proceso puede ser operado.

[1] Los orígenes de la MSR como tal se remiten al trabajo de Box y Wilson (1951), pero fue en los últimos 20 años que, debido en parte a las computadoras, esta metodología ha tenido un desarrollo considerable tanto en aspectos teóricos como en aplicaciones. Este desarrollo se refleja en varias publicaciones sobre el tema, entre las que destacan Myers y Montgomery (1995), Box y Draper (1987), Khuri y Cornell (1987) y Cornell (2002).

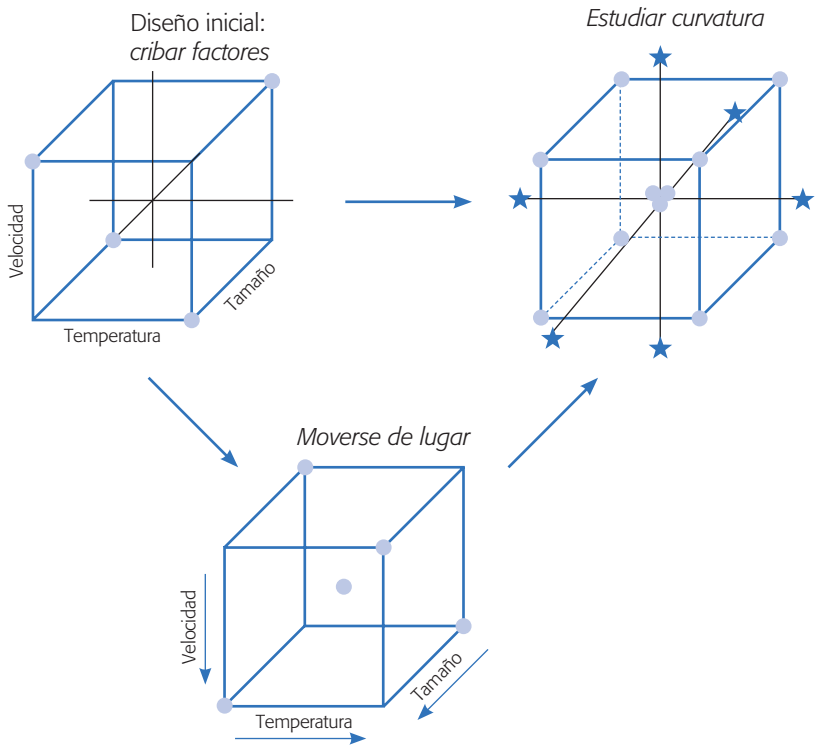


Figura 12.1 Las acciones básicas en metodología de superficie de respuesta (MSR).

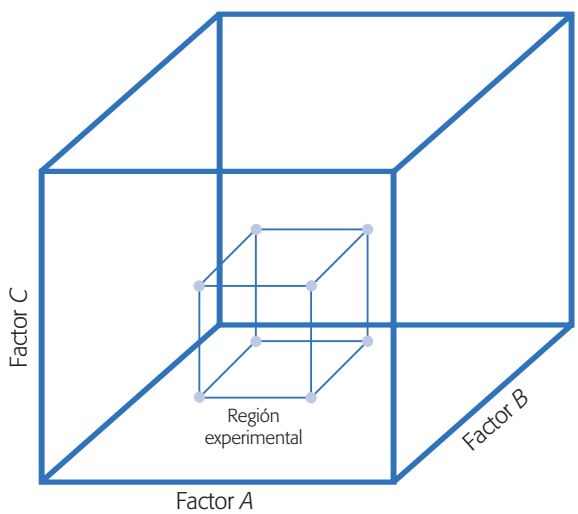


Figura 12.2 Ejemplo de regiones de operabilidad (cubo mayor) y experimental.

propuesta para el experimento inicial, y en ese caso primero será necesario acercarse a dicho punto para luego "atraparlo".

Mejor tratamiento
Mejor combinación de niveles de los factores en la cual se consideran los niveles utilizados durante el estudio experimental.

Mejor tratamiento y punto óptimo

Hasta el capítulo anterior y como conclusión de los experimentos, se encontró el *mejor tratamiento* o *mejor combinación de niveles de los factores estudiados*, y muchas

Punto óptimo
Mejor combinación de valores de los factores estudiados en la cual se considera toda la región de operabilidad.

veces éste resulta ser uno de los que se corrieron en el experimento. En particular, en diseños factoriales completos el mejor tratamiento es el "tratamiento ganador", desde el punto de vista estadístico, de entre todos los que se probaron en el estudio. En cambio, el *punto óptimo* implica que es la mejor combinación posible en toda la región de operabilidad. Así, determinar el *punto óptimo* plantea un reto más fuerte para el experimentador y requiere de una estrategia más completa, que incluye la posibilidad de realizar varios experimentos en forma secuencial y el uso de otras técnicas de análisis.

En la figura 12.3 se muestra la diferencia entre punto óptimo y mejor tratamiento. Supóngase que las *curvas de nivel* o *isolíneas* (véase capítulo 6) en esta figura representan el "verdadero comportamiento" de la respuesta, el cual tiene un punto óptimo localizado en el centro de la elipse más pequeña, que por cierto está fuera de la región experimental actual. La superficie representada en la figura se puede imaginar como una montaña y la región experimental se ubica a un costado de la cima; cada curva de nivel representa puntos sobre la montaña que tienen la misma altura. El problema es encontrar la combinación (x_{01}, x_{02}) que da por resultado el *rendimiento óptimo del proceso*.

Por otra parte, el mejor tratamiento o "tratamiento ganador" representado en la figura 12.3 es la combinación de niveles $(x_1 = -1, x_2 = 1)$, que resultaría de analizar el experimento 2^2 con punto al centro representado en la figura. Es razonable que el tratamiento ganador sea el punto experimental más cercano al verdadero punto

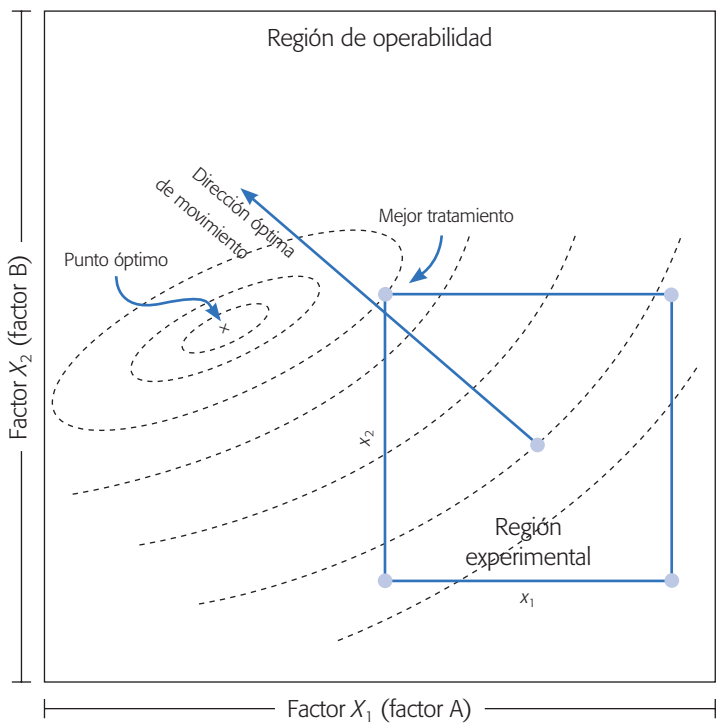


Figura 12.3 Mejor tratamiento y punto óptimo, región experimental y región de operabilidad.

óptimo, sobre todo si el experimento inicial tiene una buena región experimental. En el caso de la figura, para atrapar el óptimo es preciso desplazarse de la región actual en la mejor dirección y correr al menos otro diseño experimental que abarque al punto en cuestión y permita estudiar los efectos de curvatura pura. En la práctica, la realidad del proceso no se conoce, por lo tanto no se sabe dónde está el punto óptimo y sólo se dispone de la información obtenida en la región experimental para inferir hacia dónde se debe continuar explorando. En la figura 12.3 se supone conocida la realidad del proceso con el fin de ilustrar los conceptos, en especial la diferencia entre el mejor tratamiento y el punto óptimo.

En algunos procesos ocurrirá que el tratamiento ganador resulta ser casi tan bueno como el punto óptimo y habría que considerar si vale la pena realizar el esfuerzo de atrapar el óptimo. Por ejemplo, si el tratamiento ganador proporciona un rendimiento de 97%, ¿vale la pena realizar el esfuerzo experimental y de análisis para encontrar el punto óptimo de operación sólo por el 3% restante? La respuesta a esta pregunta depende, entre otras cosas, de cuánta ganancia en términos económicos representa 3%, y de si el proceso tiene la capacidad de dar 100% de rendimiento.

Elementos de la MSR

La metodología de superficie de respuesta implica tres aspectos: *diseño*, *modelo* y *técnica de optimización*. El diseño y el modelo se piensan al mismo tiempo, y dependen del tipo de comportamiento que se espera en la respuesta. De manera específica, el modelo puede ser de primero o segundo orden (plano o con curvatura); por ello, el tipo de diseño utilizado y el método de optimización se clasifican, según sea el caso, como de primero o segundo orden.

El aspecto *diseño* implica que para optimizar un proceso se debe aplicar el *diseño de experimentos*, en particular aquellos que sirven para ajustar un *modelo de regresión lineal múltiple* (véase capítulo 11). Más adelante se presentan algunos de estos diseños, conocidos genéricamente como *diseños para superficie de respuesta*.

El aspecto del *modelo* utiliza el *análisis de regresión lineal múltiple*, junto con sus elementos básicos que son: parámetros del modelo, modelo ajustado, significancia del modelo, prueba de falta de ajuste, residuos, predichos, intervalos de confianza para predichos y coeficiente de determinación.

Por último, el aspecto de *optimización* está formado por algunas técnicas matemáticas que sirven para que, dado un modelo ajustado, explorarlo a fin de obtener información sobre el punto óptimo. Conviene recordar técnicas como: derivadas de funciones, multiplicadores de Lagrange, operaciones con matrices, valores y vectores propios y sistemas de ecuaciones simultáneas.

En la figura 12.4 se presenta un esquema de la metodología de superficie de respuesta, donde se distinguen tres etapas en la búsqueda del punto óptimo, que son: *cribado*, *búsqueda I* o *de primer orden* y *búsqueda II* o *de segundo orden*. A continuación se describe brevemente cada una de estas etapas, más adelante se ven con detalle.

1. *Cribado*. La optimización de un proceso se inicia con esta etapa cuando tiene muchos factores (más de 6 u 8) que influyen en la variable de interés.

Diseño en MSR
Experimento apropiado basado en el conocimiento actual acerca de la posible ubicación del punto óptimo y el modelo de regresión que se quiere ajustar.

Modelo en MSR
Es la ecuación matemática que relaciona la variable de respuesta con los factores estudiados en el diseño. Por lo general es un modelo de regresión múltiple.

Optimización
Técnica matemática que sirve para extraer la información sobre el punto óptimo que tiene el modelo ajustado.

Cribado
Etapa inicial de la optimización de un proceso en la que se tienen muchos factores que pueden influir en la variable de interés.

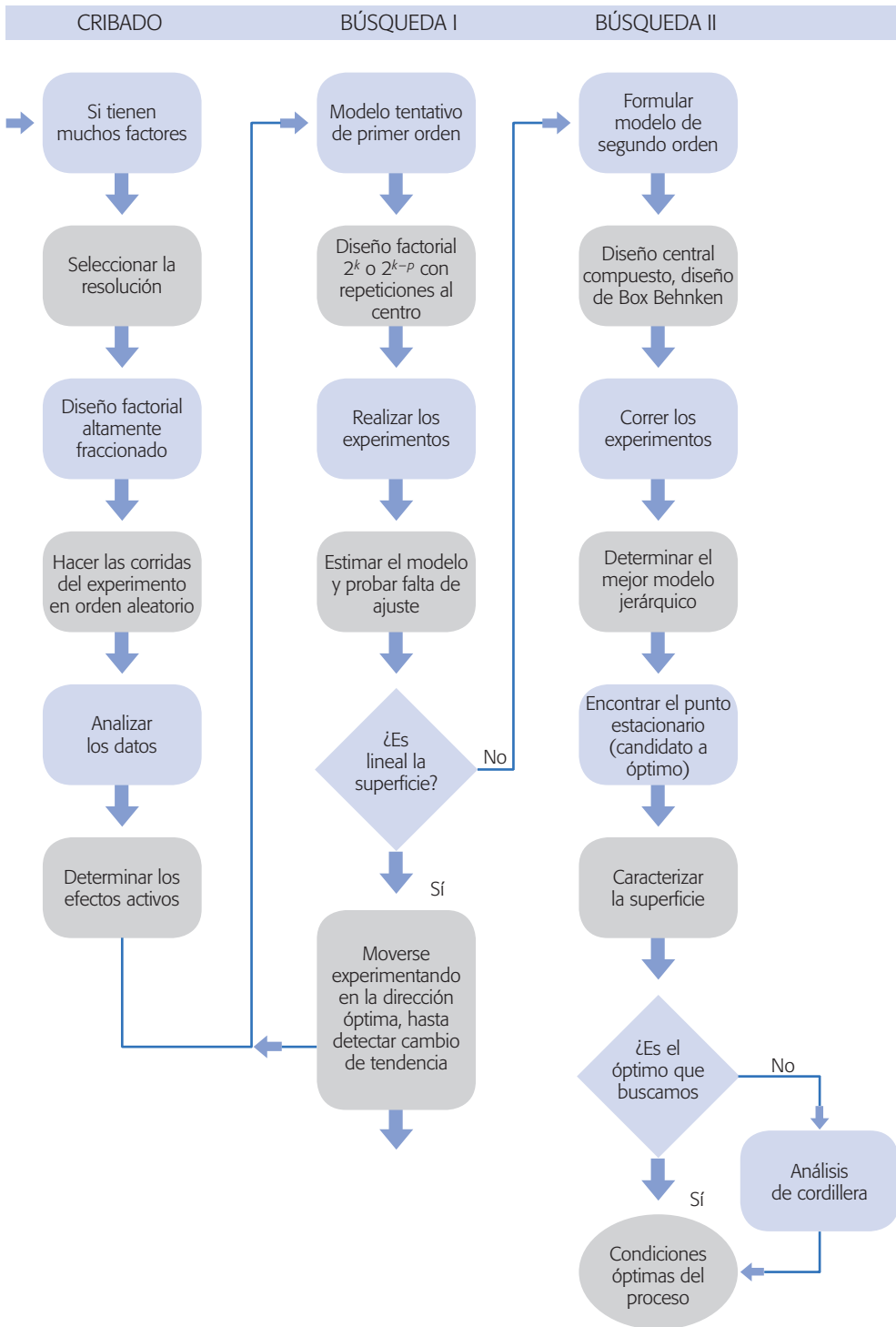

CRIBADO
BÚSQUEDA I
BÚSQUEDA II
Si tienen muchos factores
Seleccionar la resolución
Diseño factorial altamente fraccionado
Hacer las corridas del experimento en orden aleatorio
Analizar los datos
Determinar los efectos activos
Modelo tentativo de primer orden
Diseño factorial 2^k o 2^{k-p} con repeticiones al centro
Realizar los experimentos
Estimar el modelo y probar falta de ajuste
¿Es lineal la superficie?
No
Sí
Moverse experimentando en la dirección óptima, hasta detectar cambio de tendencia
Formular modelo de segundo orden
Diseño central compuesto, diseño de Box Behnken
Correr los experimentos
Determinar el mejor modelo jerárquico
Encontrar el punto estacionario (candidato a óptimo)
Caracterizar la superficie
¿Es el óptimo que buscamos
No
Sí
Análisis de cordillera
Condiciones óptimas del proceso

Figura 12.4 Esquema de los elementos de la MSR en su contexto amplio.

Por ejemplo, pensemos en una máquina que se puede manipular en 10 parámetros diferentes y que no se tiene una idea clara de cómo influye cada uno de ellos; en primer lugar es preciso correr un experimento para identificar los pocos factores que tienen mayor influencia.

2. *Búsqueda I* o *de primer orden*. Esta etapa se aplica cuando se tienen pocos factores ($k \leq 5$), y se sabe que éstos influyen en la variable de respuesta.

   En esta etapa se corre un diseño de primer orden que permita caracterizar en forma preliminar el tipo de superficie de respuesta y detectar la presencia de curvatura. Por lo general se utiliza un diseño factorial completo o fraccionado con *repeticiones al centro*.
3. *Búsqueda II* o *de segundo orden*. En el momento en que se detecta la presencia de curvatura, o bien, que la superficie es más complicada que un hiperplano, se corre o se completa un diseño de segundo orden para caracterizar mejor la superficie y modelar la curvatura. Con el modelo ajustado se determinan las condiciones óptimas de operación del proceso (véase figura 12.5).

Búsqueda de primer orden
Etapa de MSR en la que se utiliza un diseño-modelo de primer orden para caracterizar de manera preliminar la superficie y detectar curvatura.

Búsqueda de segundo orden
Etapa de MSR en la que se utiliza un diseño-modelo de segundo orden con el que se caracteriza adecuadamente la superficie de respuesta, incluyendo la curvatura.

Si la superficie no tiene curvatura y es descrita de manera adecuada por el modelo de primer orden, entonces este modelo se utiliza para moverse experimentando en la mejor dirección hasta detectar un cambio de tendencia (véase figura 12.5). En este caso se aplica de nuevo la búsqueda I. Pero si hay curvatura o la superficie es más complicada se pasa a la búsqueda II.

La metodología de superficie de respuesta se representa en la figura 12.5, en la cual se supone ya rebasada la etapa de cribado y se presentan sólo las etapas de búsquedas de primero y segundo orden (I y II), considerando dos variables de proceso. La realidad del proceso está representada por las curvas de nivel, y el punto óptimo deseado se encuentra en el centro de la superficie más pequeña, marcado con una cruz. En la práctica no se conoce *a priori* dónde se ubica el punto óptimo debido a que la realidad se desconoce, no obstante, la MSR es buena estrategia para llegar a éste.[2]

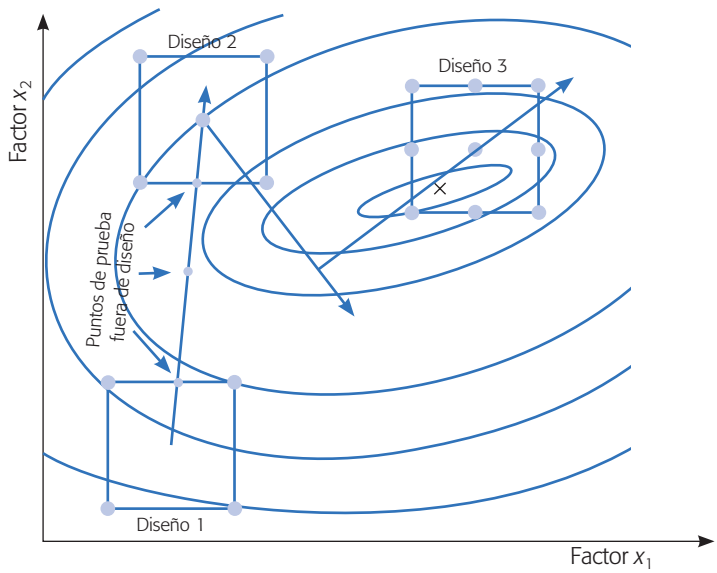


Figura 12.5 Visión gráfica de la metodología de superficie de respuesta.

[2] Aunque baja, existe la posibilidad de que la MSR lleve a un óptimo local, por lo que el experimentador debe estar consciente de ello y tener idea del potencial del proceso.

Imagínese a los tres diseños como ventanas por las que se observa la realidad desconocida del proceso; el modelo ajustado sobre cada diseño representa un aproximado a esa realidad. En el primer diseño se está lejos del punto óptimo, y el comportamiento de la superficie se modela bien con un plano y no se detecta curvatura. Con este primer modelo se encuentran puntos en la dirección de máximo ascenso para probarlos en el proceso, y se experimenta en ellos hasta que el proceso no sigue la tendencia marcada por el plano. El último punto en donde el proceso mantuvo la tendencia es el centro del diseño 2.

En el diseño 2 vuelve a ser suficiente un modelo de primer orden para modelar el comportamiento de la respuesta. Se determina la dirección óptima de movimiento, se experimenta en esa dirección hasta detectar que ya no conviene seguirla. Asimismo, se cambia de rumbo sin experimentar, al observar de dónde se partió; se determinan puntos para probar el proceso en esta nueva dirección hasta detectar un cambio en la tendencia. Entonces, se plantea un tercer diseño, cuyo punto al centro ahora sí detecta la presencia de curvatura. Se aumenta el diseño, en este caso a un 3^2 y se estima el modelo de segundo orden. Puesto que el punto óptimo se encuentra dentro de la región experimental puede atraparse determinando sus coordenadas.

Modelos

Como se explicó antes, las superficies de respuesta se caracterizan ajustando un modelo a los datos experimentales. Los modelos que se utilizan en MSR son básicamente polinomios. De esta manera, si se tienen k factores, el *modelo de primer orden* está dado por:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon \qquad \textbf{(12.1)}$$

y el *modelo de segundo orden* es:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1<j=1}^{k}\sum^{k} \beta_{ij} x_i x_j + \varepsilon \qquad \textbf{(12.2)}$$

La forma de estimar los parámetros de estos modelos y su interpretación se puede consultar en el capítulo 11.

En la figura 12.6 se muestran las gráficas para los modelos dados por las ecuaciones 12.1 y 12.2, en donde se consideran dos variables de proceso x_1 y x_2, así como diferentes valores de los parámetros. La figura 12.6*a* representa un modelo de primer orden y se observa que su superficie es un plano. En las figuras 12.6*b*, *c* y *d* se representan varios modelos de segundo orden. La forma específica que toma la superficie depende de los signos y magnitudes de los coeficientes en el modelo. En las figuras se representan las tres formas básicas, que son: *b*) superficie de máximo (montaña), *c*) superficie con mínimo (valle) y *d*) superficie con punto silla (minimax).

Para más de dos factores las superficies de respuesta no se pueden graficar completas de una sola vez porque se encuentran en cuatro dimensiones o más, pero se preserva la misma idea. Esto es, para $k > 2$ el modelo de primer orden representa un *hiperplano* y el de segundo orden constituye un *hiperelipsoide* o *hiperboloide*. Sin embargo, para $k = 3$ factores es posible graficar la superficie haciendo las tres gráficas con dos factores cada vez, con el tercero constante.

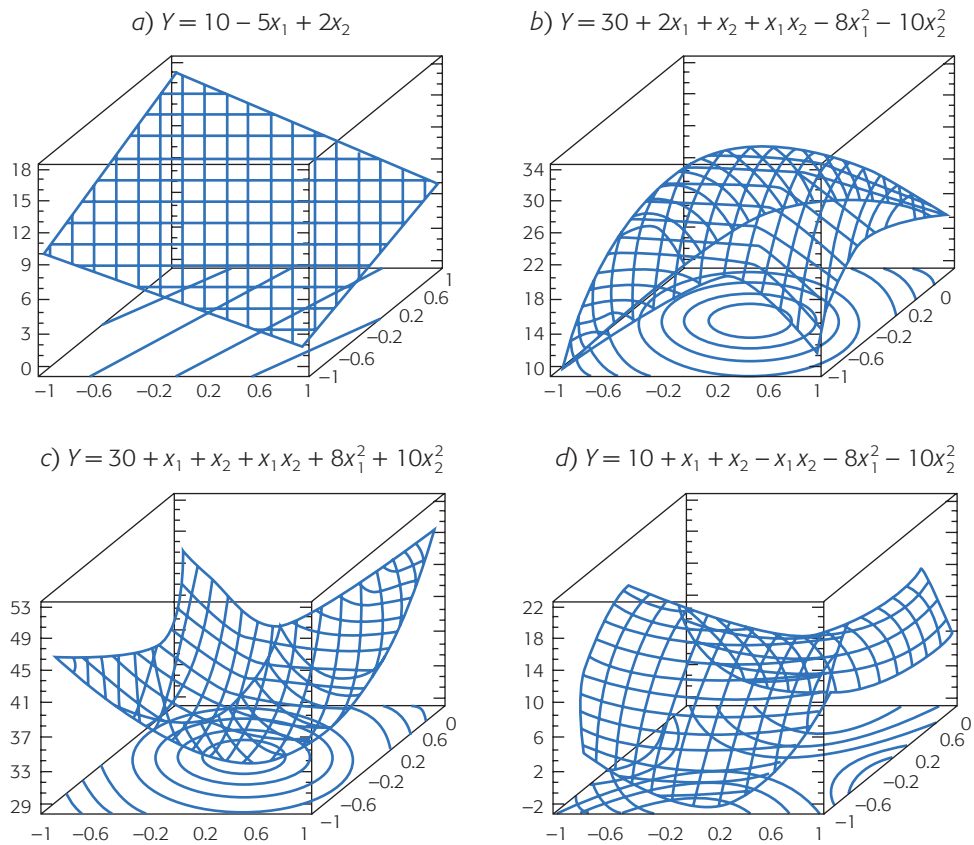


Figura 12.6 Superficies de respuesta: *a*) descrita por un modelo de primer orden; *b*), *c*) y *d*) descritas por modelos de segundo orden.

Modelos jerárquicos. Un modelo de superficie de respuesta es jerárquico si contiene todos los términos más simples que componen los términos de mayor orden que están en el modelo. Por ejemplo, el modelo $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_{12} x_1 x_2 + \hat{\beta}_{11} x_1^2$ no es jerárquico, puesto que no está el término $\hat{\beta}_2 x_2$, pero sí contiene la interacción $\hat{\beta}_{12} x_1 x_2$; para convertirlo en un modelo jerárquico habría que incluir el término $\hat{\beta}_2 x_2$ o eliminar el término de interacción $\hat{\beta}_{12} x_1 x_2$.

En superficie de respuesta se prefieren los modelos jerárquicos, ya que tienen un comportamiento más estable y suave que facilita la exploración de las superficies que representan. Esto implica que la eliminación de efectos o términos del modelo debe ser menos estricta que en análisis de varianza, cuando el objetivo es encontrar un tratamiento ganador, además de permitir que algunos términos no significativos permanezcan en el modelo para lograr la jerarquía.

Técnicas de optimización

Una vez que se tiene el modelo debidamente ajustado y validado se procede a explorar la superficie descrita por el modelo para encontrar la combinación de niveles en los factores que dan por resultado un valor óptimo de la respuesta, o bien, para deter-

minar la dirección óptima de movimiento en la que se debe experimentar en el futuro. Si el modelo no explica un mínimo de 70% del comportamiento de la respuesta, en términos del R^2_{aj}, no se recomienda utilizarlo para fines de optimización porque su calidad de predicción es mala. En adelante supondremos niveles codificados para los factores (–1, +1), lo cual facilita las interpretaciones y los cálculos. Por lo que siempre que se encuentren las condiciones óptimas o la dirección de experimentación futura, primero se hará en condiciones codificadas y después eso se debe traducir a condiciones o niveles reales. Aunque el uso de un software puede evitar el uso de códigos.

La técnica de optimización a utilizar depende del tipo de modelo ajustado y existen básicamente tres métodos, que son:

1. Escalamiento ascendente (o descendente)
2. Análisis canónico
3. Análisis de cordillera

El *escalamiento ascendente* es para el modelo de primer orden y las otras dos técnicas son para el modelo de segundo orden. A continuación se describen cada uno de estos métodos.

Escalamiento ascendente (descendente)

Escalamiento ascendente
Técnica que sirve para determinar puntos (tratamientos) por experimentar que están ubicados sobre la dirección de ascenso máximo a partir del centro del diseño inicial.

Cuando la variable de respuesta de interés es del tipo: mientras más grande es mejor, se tiene un *escalamiento ascendente*; pero si lo que interesa es: mientras más pequeña mejor, se trata de *escalamiento descendente*.

De aquí en adelante, diremos simplemente escalamiento ascendente, en lugar de "escalamiento ascendente (descendente)", puesto que el escalamiento descendente se convierte en ascendente al cambiar los signos de los términos del modelo ajustado. Cuando la respuesta es del tipo: el valor nominal es lo mejor, el problema es localizar la curva de nivel específica que tenga la altura o valor requerido de la variable de respuesta. En este caso, cada punto sobre la curva de nivel es una solución, y de todos ellos se elige el de menor variabilidad y/o menor costo.

La técnica de *optimización de escalamiento* se aplica cuando, de acuerdo con la valoración inicial, se cree que se está lejos de la condición óptima, por lo que será necesario explorar una región de experimentación inicial y a partir de ésta determinar una dirección en la cual experimentar fuera de la región inicial.

Así, a partir del conocimiento que ya se tiene del problema es preciso seleccionar los niveles de los factores para determinar la región de exploración. A continuación se corre un diseño de primer orden (típicamente un diseño 2^k completo o fraccionado con puntos al centro) para explorar la región experimental determinada antes. Se analizan con detalle los resultados y se ajusta un modelo de primer orden con niveles codificados. Si éste explica satisfactoriamente la variabilidad observada es necesario continuar como se indica más adelante, de lo contrario, investigar a qué se debe la falta de ajuste: ¿Mucha variabilidad? ¿Curvatura? ¿Región más complicada? y proceder en consecuencia. Con el siguiente ejemplo vamos a ilustrar la metodología.

Ejemplo 12.1

Consideremos el problema de diseñar un helicóptero de papel que se mantenga en el aire el mayor tiempo posible (véase Box y Liu, 1999). La manera de construir un helicóptero con una hoja de papel tamaño carta se ilustra en el diagrama de la figura 12.7. Son muchos los factores que influyen en el tiempo de vuelo, por lo que se decide comenzar por la etapa de cribado (véase figura 12.4), en la cual, con un diseño factorial fraccionado, se busca determinar los pocos factores que en realidad influyen sobre el tiempo de vuelo. Los factores con sus correspondientes niveles considerados en este experimento se enlistan a continuación:

Factor	−1	+1
A: Tipo de papel	regular	bond
B: Longitud de alas	3.00 pulg.	4.75 pulg.
C: Longitud del cuerpo	3.00 pulg.	4.75 pulg.
D: Ancho del cuerpo	1.25 pulg.	2.00 pulg.
E: Cuerpo doblado	no	sí
F: Cuerpo con adhesivo	no	sí
G: Clip al cuerpo	no	sí
H: Alas con adhesivo	no	sí

Una vez que se doblan y se pliegan las partes laterales del cuerpo, el factor *E* se refiere a que el cuerpo se dobla hacia la parte superior.

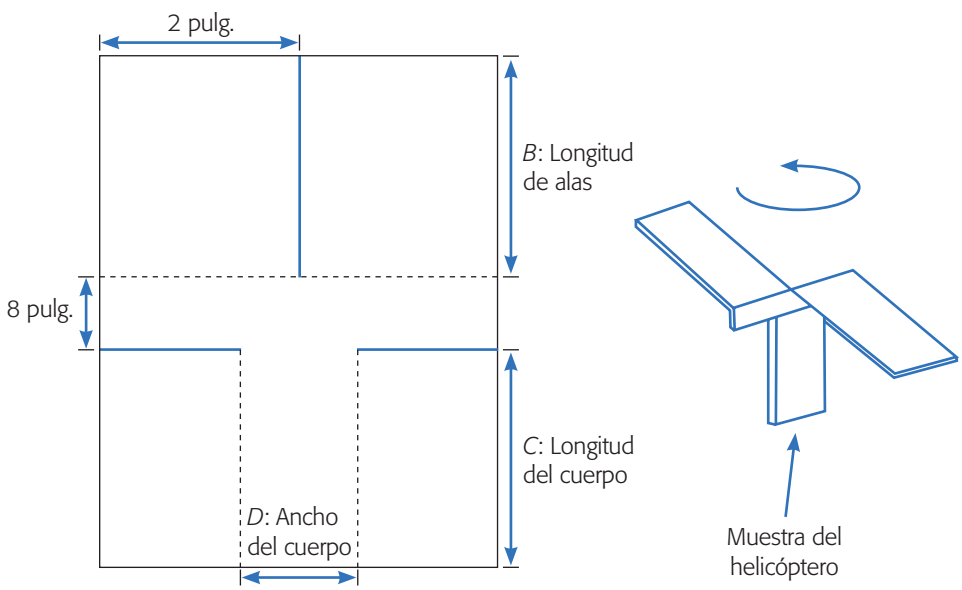


Figura 12.7 Construcción de un helicóptero de papel y sus factores, con _____ doblar y ——— cortar.

Note que hay factores con los que se decidió no experimentar, como son el ancho de las alas (se ha fijado en 2 pulg.) y el ancho del cuerpo medio (se fija en 0.8 pulg.). La variable de respuesta es el tiempo de vuelo en centésimas de segundo de cada helicóptero al dejarlo caer desde el techo (2.70 metros). Se utiliza un diseño factorial fraccionado 2^{8-4}_{IV}. Cada uno de los 16 helicópteros realizados se lanza cuatro veces para estimar mejor el tiempo en el aire. Los tratamientos y los promedios de los cuatro tiempos de vuelo se muestran en la tabla 12.1. Se cuida que cada corrida sea en igualdad de circunstancias para evitar sesgos (los lanza la misma persona, lo trata de hacer de la misma manera, etc.). Aplicando lo visto en el capítulo 8 se hace un análisis de los datos y se obtiene el mejor análisis de varianza de la tabla 12.1. Como se aprecia, sólo afectan estadísticamente los factores continuos *B*, *C* y *D* listados arriba, además del factor categórico *G* (clip). El modelo de regresión ajustado considerando sólo los factores continuos, en niveles codificados (–1, +1), resulta ser:

$$\hat{Y} = 223 + 28x_2 - 13x_3 - 8x_4$$

Aquí, *B* es x_2, *C* es x_3 y *D* corresponde a x_4. El R^2 en la tabla 12.1 indica que se tiene un modelo que explica de manera satisfactoria la variación en $\hat{Y}$.

Con el modelo ajustado y depurado se determina la dirección óptima de movimiento a partir del centro del diseño. Durante esta etapa los factores que no influyeron se mantienen en sus niveles más económicos. La magnitud de F_0 en el ANOVA indica qué factor fue más importante, y los signos de los coeficientes indican si el nivel del factor correspondiente se debe incrementar o disminuir para llevar a un mejor valor a *Y*.

Para obtener la trayectoria es necesario decidir una longitud de paso en unidades codificadas y reales. Se recomienda utilizar un paso de movimiento unitario (en unidades codificadas) en el factor con mayor influencia, con lo que se asegura que los pasos en los factores restantes serán de menor amplitud y proporcionales a sus coeficientes.

Un paso unitario equivale a moverse en el factor correspondiente a intervalos de un medio de su rango de prueba en unidades originales. Una longitud de paso mayor se considera agresiva y puede tener riesgos, mientras que una longitud de paso menor se considera una forma conservadora de proceder, que algunos casos que se está experimentando a nivel proceso puede ser una buena opción (véase ejemplo 12.4).

En el caso concreto del modelo para el helicóptero, la dirección óptima es "por cada 28 unidades que se incrementen al factor x_2, deben reducirse 13 unidades al factor x_3 y disminuirse en 8 al factor x_4". Como el factor de mayor influencia es x_2, entonces, de acuerdo con los niveles utilizados en el diseño, se tiene que una unidad codificada en el factor x_2 equivale a (4.75 – 3.00)/2 = 0.875 pulgadas, que es la mitad del rango experimental utilizado. Es decir, el paso codificado para el factor x_2 es $\Delta x_2 = 1$ y en unidades originales es $\Delta Z_2 = 0.875$ pulgadas. Los pasos en unidades codificadas para los factores restantes, en este caso x_3 y x_4, se determinan dividiendo sus coeficientes en el modelo entre el coeficiente del factor base x_2 y multiplicando por el paso $\Delta x_2 = 1$. Al realizar los cálculos se tiene que:

$$\Delta x_3 = -13/28 = -0.46$$

Tabla 12.1 Resultados de helicópteros construidos y análisis de varianza.

Tratamiento	A	B	C	D	E	F	G	H	$\bar{y}$
1	–1	–1	–1	–1	–1	–1	–1	–1	236
2	1	–1	–1	–1	–1	1	1	1	185
3	–1	1	–1	–1	1	–1	1	1	259
4	1	1	–1	–1	1	1	–1	–1	318
5	–1	–1	1	–1	1	1	1	–1	180
6	1	–1	1	–1	1	–1	–1	1	195
7	–1	1	1	–1	–1	1	–1	1	246
8	1	1	1	–1	–1	–1	1	–1	229
9	–1	–1	–1	1	1	1	–1	1	196
10	1	–1	–1	1	1	–1	1	–1	203
11	–1	1	–1	1	–1	1	1	–1	230
12	1	1	–1	1	–1	–1	–1	1	261
13	–1	–1	1	1	–1	–1	1	1	168
14	1	–1	1	1	–1	1	–1	–1	197
15	–1	1	1	1	1	–1	–1	–1	220
16	1	1	1	1	1	1	1	1	241

Análisis de varianza para tiempo de vuelo promedio

FV	SC	GI	CM	F_0	Valor-*p*
B	12 321.0	1	12 321.0	36.87	0.0001
C	2 809.0	1	2 809.0	8.41	0.0145
D	1 089.0	1	1 089.0	3.26	0.0984
G	1 892.3	1	1 892.3	5.66	0.0365
Error	3 675.8	11	334.2		
Total	21 787.0	15			
$R^2 = 83.1\%$, $R^2_{(aj)} = 77.0\%$					

$$\Delta x_4 = -8/28 = -0.28$$

ambas son unidades codificadas. Mediante la regla de tres se encuentran los pasos en unidades originales para estos dos factores. Para el factor x_3, si 1 unidad codificada para este factor equivale a 0.875 pulgadas, entonces 0.46 unidades codificadas equivalen a $0.875 \times 0.46/1 = 0.40$ pulgadas. De manera similar, para el factor x_4, si una unidad codificada equivale a $(2 - 1.25)/2 = 0.375$ pulgadas, entonces 0.28 unidades codificadas equivalen a $0.375 \times 0.28/1 = 0.10$ pulgadas. En resumen, encontramos los siguientes pasos: $\Delta x_2 = 1$, $\Delta x_3 = -0.46$, $\Delta x_4 = -0.28$, lo cual en unidades originales (pulgadas) equivale a $\Delta Z_2 = 0.875$, $\Delta Z_3 = -0.40$ y $\Delta Z_4 = -0.10$.

Con la información generada hasta este momento se procede a construir la tabla 12.2 de escalamiento ascendente, que debe contener al menos cuatro puntos

sobre la dirección óptima, a partir del centro del diseño y expresados en unidades codificadas y originales. Los valores codificados y reales del paso 1 se obtienen al sumar la longitud de paso al centro del diseño, y así se suman tantos pasos como puntos se necesiten. Note que al sumar un paso negativo en realidad se está restando, por eso los niveles de los factores x_3 y x_4 van decreciendo.

A continuación, de acuerdo con la tabla de escalamiento se hacen las corridas del proceso sobre la dirección óptima, que en este problema significa diseñar los helicópteros. Igual que antes, se decidió que el tamaño de prueba fuera de cuatro lanzamientos en cada condición debido a que se tiene un proceso rápido; con los cuatro lanzamientos se obtuvo un tiempo promedio de vuelo en cada condición. Note que repetir un lanzamiento es volver a medir y no es una réplica, ya que ésta implicaría hacer un helicóptero con las mismas dimensiones teóricas. Los valores encontrados en las corridas de la tabla 12.2 fueron: 236 en el centro, 311 en el paso 1, 356 en el paso 2 y 321 en el paso 3. En ese momento se detecta un cambio en la tendencia ascendente del tiempo promedio de vuelo; del paso 2 al paso 3 el tiempo observado fue menor. Si graficamos el número de paso contra el tiempo de vuelo se obtiene la figura 12.8. Cuando en esta gráfica se detecta claramente que *Y* ha dejado de incrementarse, entonces ya no debe seguirse la dirección de búsqueda dada por la tabla de escalamiento. En el caso del helicóptero, en el tercer paso claramente no se incrementó el tiempo de vuelo, por lo que ya no deben hacerse los helicópteros del paso 4 y 5.

Nótese que con apenas 19 pruebas se consiguió identificar los factores críticos y tener un helicóptero que vuela durante 356 centésimas de segundo, que es el doble del peor de los helicópteros de la tabla 12.1. Ahora, lo que sigue es determinar el centro de la nueva región experimental y volver a aplicar la metodología.

De acuerdo con lo anterior, el punto del paso 2 fue el último donde se mantuvo la tendencia ascendente del tiempo de vuelo, por lo tanto, éste puede ser el centro de una nueva región de experimentación. Como ahora se tienen sólo tres factores es recomendable utilizar un diseño que permita destacar la presencia de efectos de curvatura pura, por ejemplo un factorial 2^3 + centro.

Con los datos obtenidos se ajusta un modelo de primer orden y se estima la curvatura. Si ya se está cerca del óptimo, entonces la curvatura será significativa y/o el modelo de primer orden no describirá de manera satisfactoria la región que se explora. En ese caso será necesario completar un diseño de segundo orden que

Tabla 12.2 Escalamiento ascendente para el tiempo de vuelo.

	Códigos			Niveles reales			
	x_2	x_3	x_4	Z_2	Z_3	Z_4	Y
Centro	**0**	**0**	**0**	**3.875**	**3.875**	**1.625**	**–**
Longitud de paso	**1**	**–0.46**	**–0.28**	**0.875**	**–0.40**	**–0.10**	
Paso 1	1	–0.46	–0.28	4.75	3.475	1.525	–
Paso 2	2	–0.92	–0.56	5.625	3.075	1.425	–
Paso 3	3	–1.38	–0.84	6.500	2.675	1.325	–
Paso 4	4	–1.84	–1.12	7.375	2.275	1.225	–
Paso 5	5	–2.30	–1.40	8.250	1.875	1.125	–

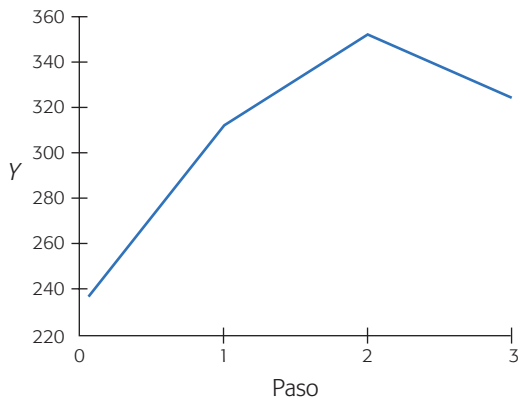


Figura 12.8 Gráfica de número de paso contra el tiempo de vuelo para el helicóptero.

permita ajustar un modelo de segundo orden para, con base en éste, encontrar las condiciones de los tres factores de la tabla 12.2 que maximizan el tiempo de vuelo de los helicópteros de papel.

Análisis canónico

Se aplica un diseño de segundo orden cuando se quiere explorar con más amplitud una región experimental y/o cuando se espera que el punto óptimo ya esté cerca (probablemente dentro de la región experimental). El *análisis canónico* es una de las técnicas para analizar el modelo de segundo orden y consiste en los siguientes pasos:

Análisis canónico
Técnica empleada para caracterizar la superficie de segundo orden: las coordenadas del punto estacionario, el tipo de punto y la orientación de la superficie.

1. A partir del conocimiento que ya se tiene del problema, seleccionar los niveles de los factores para determinar la región de exploración.
2. Correr un diseño de segundo orden (un diseño de composición central, por ejemplo) para explorar la región experimental determinada antes.
3. Ajustar un modelo de segundo orden con niveles codificados. Si éste explica bien la variabilidad observada continuar al siguiente paso; de lo contrario, investigar por qué la falta de ajuste (¿mucha variabilidad?, ¿región más complicada?) y proceder en consecuencia.
4. Encontrar las coordenadas del punto estacionario.
5. Expresar el modelo ajustado en su forma canónica. El análisis canónico consiste en reescribir el modelo ajustado de segundo orden en su *forma canónica*, es decir, se expresa en términos de nuevas variables llamadas *variables canónicas*, las cuales son transformaciones de las variables codificadas. La ventaja es que la ecuación canónica proporciona información a simple vista sobre el tipo de superficie que se está observando y sobre su forma.
6. Evidenciar la relación entre las variables canónicas y las variables codificadas.

En la práctica, si se cuenta con un software adecuado no necesariamente se siguen los últimos tres pasos del análisis canónico. La mejor estrategia será encontrar, primero los coeficientes de la ecuación canónica que indican el tipo de superficie observada y sólo si ésta es del tipo que interesa (por ejemplo máximo), entonces se

procede a localizar las coordenadas del punto estacionario. Si la superficie encontrada no es del tipo deseado se sigue el análisis de cordillera descrito en la siguiente subsección. Sin embargo, primero veamos cómo determinar el punto estacionario, dado que interviene en el término independiente de la ecuación canónica.

Punto estacionario
Es el punto en el espacio de factores, sobre el cual el plano tangente a la superficie tiene pendiente igual a cero y es un candidato a óptimo.

Determinación del punto estacionario (candidato a óptimo). El *punto estacionario* es el punto $(x_{10}, x_{20}, \ldots, x_{k0})$ en el espacio de factores, sobre el cual el plano tangente a la superficie tiene pendiente igual a cero. Por ejemplo, si la superficie tiene un máximo, el punto estacionario es justo el punto donde se ubica ese máximo. De aquí que el punto estacionario sea un candidato natural a punto óptimo, que resulta "electo" sólo cuando es del tipo que interesa y se encuentra dentro de la región experimental. Podría pasar que aunque se esté buscando un máximo, el punto estacionario sea un mínimo o punto silla, en cuyo caso evidentemente no se trataría del óptimo buscado.

Suponga que ya se realizaron los tres primeros pasos de un análisis canónico, y que por lo tanto ya se tiene ajustado un modelo de segundo orden:

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \sum_{i=1}^{k} \hat{\beta}_{ii} x_i^2 + \sum_{i=1<j=1}^{k} \sum \hat{\beta}_{ij} x_i x_j$$

para el cual se quiere encontrar su punto estacionario (donde la derivada es igual a cero). El punto se localiza derivando al modelo con respecto a cada variable x_i, igualando a cero y resolviendo en forma simultánea todas las ecuaciones. Todo esto se facilita si el modelo se reescribe en notación matricial como:

$$\hat{Y} = \hat{\beta}_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x} \qquad \textbf{(12.3)}$$

donde $\mathbf{x}' = (x_1, x_2, \ldots, x_k)$ es cualquier punto en la región de operabilidad del proceso, en unidades codificadas; el vector $\mathbf{b}$ son los coeficientes de la parte lineal (efectos principales) del modelo y la matriz $\mathbf{B}$ son los coeficientes de las interacciones y de los términos cuadráticos puros. Esto es:

$$b = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}; \quad B = \begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & \hat{\beta}_{13}/2 & \cdots & \hat{\beta}_{1k}/2 \\ \hat{\beta}_{12}/2 & \hat{\beta}_{22} & \hat{\beta}_{23}/2 & \cdots & \hat{\beta}_{2k}/2 \\ \hat{\beta}_{13}/2 & \hat{\beta}_{23}/2 & \hat{\beta}_{33} & \cdots & \hat{\beta}_{3k}/2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\beta}_{1k}/2 & \hat{\beta}_{2k}/2 & \hat{\beta}_{3k}/2 & \cdots & \hat{\beta}_{kk} \end{pmatrix} \qquad \textbf{(12.4)}$$

Derivando el modelo dado por (12.3) con respecto al vector $\mathbf{x}$ e igualando a cero se obtiene:

$$\frac{\partial \hat{Y}}{\partial \mathbf{x}} = \mathbf{b} + 2\mathbf{B}\mathbf{x} = 0$$

Resolviendo para $\mathbf{x}$ se llega a que el punto estacionario está dado por:

$$\mathbf{x}_0 = \frac{-\mathbf{B}^{-1}\mathbf{b}}{2} \qquad \textbf{(12.5)}$$

donde $\mathbf{B}^{-1}$ es la inversa de la matriz $\mathbf{B}$.

Ejemplo 12.2

Un problema que se tiene en el proceso de fabricación de cámaras de un solo uso es la adherencia de fibras a la pared interior de la cámara, que al hacerse visibles en las fotografías causan insatisfacción en el consumidor final. El problema se abordó atacando las fuentes de contaminación y mediante la operación de un equipo de limpieza de las cámaras. En este tipo de equipo, que es como una especie de aspiradora, se va a utilizar diseño de experimentos para buscar optimizar su funcionamiento. Un primer problema que se presentó fue cómo experimentar con las cámaras si no se sabe qué nivel de contaminación trae cada una. Esto podría sesgar los resultados del experimento, salvo que se realizara un número excesivo de pruebas (limpieza de cámaras) para contrarrestar la variación del número de fibras en el interior. Se consideró que una mejor alternativa era preparar previamente las cámaras con cierto nivel de contaminación similar al observado en el proceso. Para ello se encontró la manera de sembrar 16 fibras en el interior de las cámaras a un nivel de adherencia realista.

El equipo de limpieza tiene tres factores de control: presión de aire (x_1), tiempo de aplicación (x_2) y presión de vacío, pero por limitaciones para manipular la presión de vacío, ésta se tuvo que descartar del estudio fijándola en su nivel alto, por considerar que éste favorece la eliminación de las fibras. Para estudiar los otros dos factores se utilizó un diseño factoria1 2^2 con dos repeticiones al centro y replicado tres veces. Los datos se muestran en la tabla 12.3 y representan el número de fibras que permanecen en el interior de la cámara después de aplicar el tratamiento. Las unidades de presión son psi y segundos la del tiempo.

Se analizan los datos (tabla 12.4) y se detecta una fuerte presencia de efectos de curvatura pura, al resultar significativa la prueba de falta de ajuste *(lack of fit test*, véase capítulo 11). Esto implica que alguno o ambos términos cuadráticos puros influyen fuertemente sobre *Y*.

Para investigar los efectos cuadráticos puros x_1^2 y x_2^2 se decide aumentar el experimento inicial para convertirlo en un diseño de composición central con puntos axiales en las caras ($\alpha = 1$, ver siguiente sección). Los datos originales y los obtenidos en los puntos estrella o axiales en cada una de las tres réplicas se muestran en la tabla 12.5. Se decide realizar una corrida adicional al centro en cada réplica. El ANOVA se muestra en la tabla 12.6; se está considerando un efecto de bloque debido a que los tratamientos de los puntos axiales fueron corridos en otro momento. La interacción no resulta significativa, por lo que se eliminan. El modelo final está dado por:

Tabla 12.3 Experimento en cámaras de un solo uso.

Unidades codificadas		Unidades originales		*Rep* 1	*Rep* 2	*Rep* 3
A: x_1	*B*: x_2	Presión (psi)	Tiempo (seg)			
–1	–1	10	1	3	4	4
1	–1	30	1	2	4	3
–1	1	10	3	7	8	8
1	1	30	3	7	7	6
0	0	20	2	2	2	1
0	0	20	2	1	1	2

Tabla 12.4 ANOVA para datos de la tabla 12.3.

FV	SC	GL	CM	F_0	Valor-*p*
A: Presión	2.08	1	2.08	4.92	0.0449
B: Tiempo	44.08	1	44.08	104.20	0.0000
Falta de ajuste	56.33	2	28.17	66.58	0.0000
Error puro	5.50	13	0.42		
Total	108.00	17			

$$\hat{Y} = 2.08 - 0.83x_1 + 1.5x_2 + 1.78x_1^2 + 1.78x_2^2 \qquad \textbf{(12.6)}$$

El R^2_{aj} es satisfactorio. Las gráficas de superficie y de contornos del modelo ajustado se presentan en la figura 12.9. En ésta se observa que la superficie de respuesta tiene un mínimo dentro de la región experimental, y era precisamente un mínimo lo que se buscaba puesto que la respuesta *Y* son las fibras que permanecen en el interior de la cámara después de aplicar el tratamiento. Utilizando una regla sobre la gráfica de contornos es posible obtener las coordenadas en unidades codificadas del punto óptimo dadas por $x' = (x_{10}, x_{20}) = (0.22, -0.41)$. Estas coordenadas del punto estacionario se pueden decodificar utilizando la relación:

$$Z_i = \frac{x_i(Z_H - Z_L) + (Z_H + Z_L)}{2} \qquad \textbf{(12.7)}$$

con cada una de las variables, donde Z_H y Z_L son los niveles alto (+1) y bajo (–1) en unidades originales, dados en la tabla 12.3. Así, $Z_{1H} = 30$, $Z_{1L} = 10$, $Z_{2H} = 3$ y $Z_{2L} = 1$. Aplicando la relación (12.7) primero al factor presión (x_1) se obtiene,

$$Z_{10} = \frac{0.22(30-10) + (30+10)}{2} = 22.2 \text{ libras/pulg}^2$$

que es la coordenada del punto óptimo de la presión expresada en unidades de proceso. De manera similar, para el factor tiempo (x_2) se llega a que:

$$Z_{20} = \frac{-0.41(3-1) + (3+1)}{2} = 1.59 \text{ segundos}$$

es el tiempo óptimo. Ésta es la solución gráfica en la cual se aprovecha el hecho de que se estudiaron sólo dos factores. Normalmente aquí acabaría el análisis puesto que se ha encontrado el óptimo deseado. Pero enseguida vemos algunos detalles matemáticos que son útiles para situaciones de optimización más complicadas.

Tipos de superficie y ecuación canónica. En algunos problemas de optimización, en especial con diversas variables el tipo de superficie de respuesta no es tan claro. Por ello se utiliza la *ecuación canónica*, que es otra forma de escribir el modelo de segundo orden dado por (12.3) y (12.6). La *forma canónica* del modelo de segundo orden está dada por:

$$\hat{Y} = \hat{Y}_0 + \lambda_1 w_1^2 + \lambda_2 w_2^2 + \cdots + \lambda_k w_k^2 \qquad \textbf{(12.8)}$$

Tabla 12.5 Diseño central compuesto para el número de fibras.

Presión	Tiempo	Rep. 1	Rep. 2	Rep. 3
–1	–1	3	4	4
1	–1	2	4	3
–1	1	7	8	8
1	1	7	7	6
–1	0	6	7	6
1	0	4	2	3
0	–1	4	3	5
0	1	6	5	5
0	0	2	2	1
0	0	1	1	2
0	0	3	2	0

Tabla 12.6 Análisis de varianza para datos de tabla 12.5.

FV	*SC*	*Gl*	*CM*	F_0	Valor-*p*
x_1	12.5	1	12.5	10.61	0.0031
x_2	40.5	1	40.5	34.37	0.0000
$x_1{*}x_1$	29.7	1	29.7	25.19	0.0000
$x_1{*}x_2$	0.083	1	0.083	0.07	0.7924
$x_2{*}x_2$	29.7	1	29.7	25.19	0.0000
Bloque	9.5	1	9.5	8.08	0.0086
Error	30.6	26	1.178		
Total	158.97	32			
$R^2 = 80.7\%,\ R^2_{(aj)} = 77.12\%$					

donde $\hat{Y}_0$ es el valor predicho por el modelo sobre el punto estacionario; las ω_i son nuevas variables independientes entre sí, transformadas de las $x_i (i = 1, 2,\ldots, k)$, que se llaman *variables canónicas*. En términos geométricos la ecuación canónica es una *rotación* de los ejes coordenados del modelo original, seguida de una *traslación* de los mismos (figura 12.10) a un nuevo origen. El origen de los nuevos ejes se coloca sobre el punto estacionario, y los ejes de las variables ω_i están orientados en el sentido de la cordillera de la superficie: ω_1 corre en la dirección de la cordillera más abrupta, o sea, con declive máximo; ω_2 va en dirección perpendicular al eje de la primera variable explicando la siguiente dirección más abrupta; ω_3 explica la tercera dirección en importancia, y así sucesivamente hasta ω_k.

Los coeficientes λ de la ecuación canónica, que más adelante veremos cómo obtenerlos, son los *valores propios* o *característicos* de la matriz **B** definida en (12.4) y sus signos determinan el tipo de punto estacionario que se encontró de acuerdo con la siguiente regla:

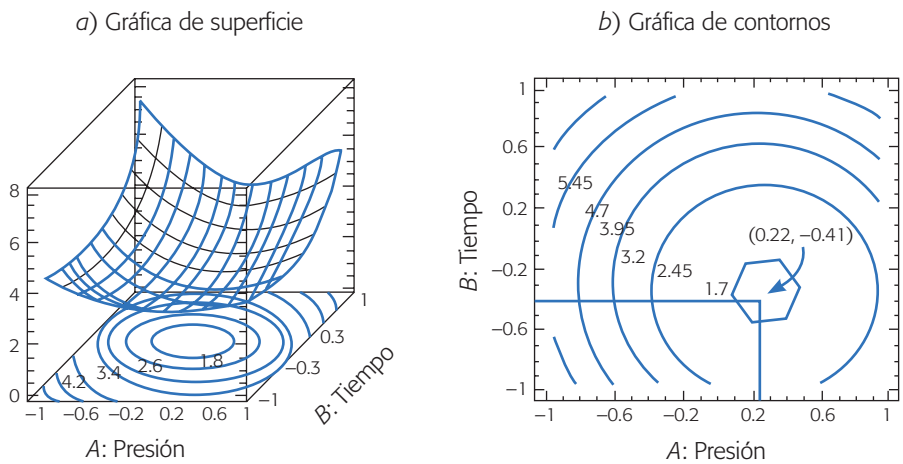


Figura 12.9 Gráficas de superficie y de contornos para modelo de (12.6).

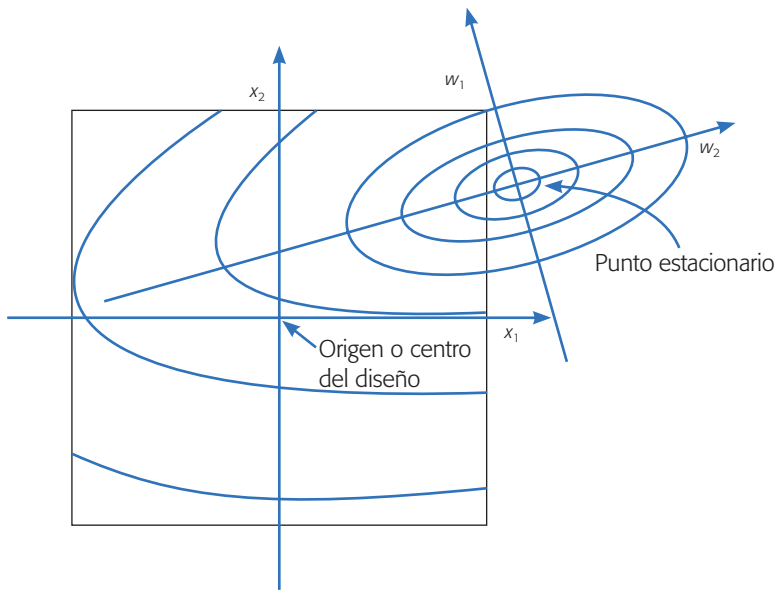


Figura 12.10 Significado gráfico de la ecuación canónica para dos factores.

1. Si λ_i es positivo para toda i, es un MÍNIMO.
2. Si λ_i es negativo para toda i, es un MÁXIMO.
3. Si hay ambos signos es un PUNTO SILLA.

Los tipos de superficie que pueden resultar se representan en la figura 12.11 para $k = 2$ factores. Además de superficies con *máximo*, *mínimo* o *punto silla*, apare-

cen otros dos casos que se derivan de éstos y que dependen de las magnitudes de los coeficientes λ_i: la *cresta ascendente* (*descendente*) ocurre cuando se tiene un máximo (o mínimo), pero el punto estacionario cae fuera de la región experimental y lo que se observa entonces en la figura es un "costado de la montaña". Cabe hacer notar que la forma de cresta ascendente también puede aparecer (aunque es más improbable) cuando la región experimental se ubica justo sobre una de las cuatro familias de parábolas que componen al punto silla. La *cordillera estacionaria* es un caso límite, ya sea del máximo o mínimo o del punto silla, y ocurre cuando uno de los valores propios se aproxima a cero.

Cresta ascendente
Es cuando se tiene un máximo, pero éste cae fuera de la región experimental.

Cordillera estacionaria
Superficie que tiene una infinidad de máximos o mínimos, representados por una curva de nivel lineal.

Así, para saber qué tipo de superficie se tiene es necesario obtener los valores propios de la matriz **B**. Los valores propios, también conocidos como valores característicos o *eigenvalores* cumplen con la relación:

$$\mathbf{B}\mathbf{m}_i = \lambda_i \mathbf{m}_i$$

donde $\mathbf{m}_i$ es un vector propio asociado al valor propio λ_i. Para obtener los valores propios se resuelve el polinomio en términos de λ, que resulta de resolver la ecuación *determinante*,

$$|\mathbf{B} - \lambda \mathbf{I}| = 0$$

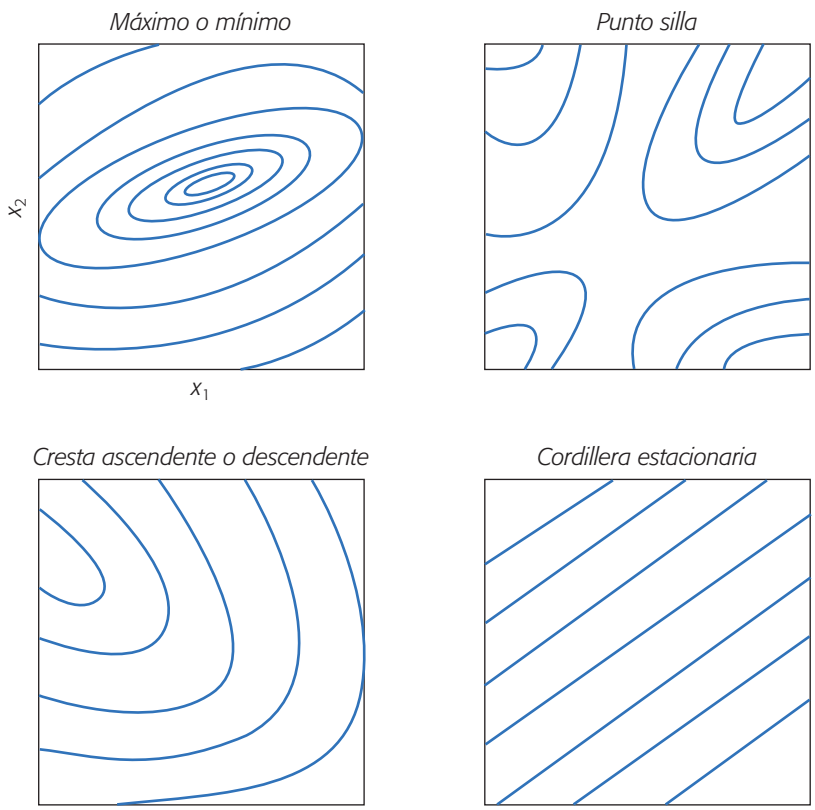


Figura 12.11 Ejemplos de los tipos de superficie de respuesta.

donde **I** es la matriz identidad de dimensión *k*. Se recomienda apoyarse en un sistema computacional para obtener los valores propios.

Ejemplo 12.3 (Continuación)

Para obtener la solución analítica del problema de las fibras en cámaras de un solo uso se utiliza el análisis canónico. Del modelo de la ecuación 12.6 se tiene que

$$\mathbf{b} = \begin{bmatrix} -0.83 \\ 1.50 \end{bmatrix}; \ \mathbf{B} = \begin{bmatrix} 1.78 & 0 \\ 0 & 1.78 \end{bmatrix}$$

Los valores y vectores propios de la matriz **B** se obtienen de resolver la ecuación $P(\lambda)$ dada por,

$$P(\lambda) = |\mathbf{B} - \lambda\mathbf{I})| = \left| \begin{pmatrix} 1.78 & 0 \\ 0 & 1.78 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = 0$$

$$\left| \begin{pmatrix} 1.78 - \lambda & 0 \\ 0 & 1.78 - \lambda \end{pmatrix} \right| = (1.78 - \lambda)^2 = 0$$

cuya solución es $\lambda_1 = 1.78$ y $\lambda_2 = 1.78$. Como ambos valores propios son positivos, el punto estacionario es un mínimo, y al ser de la misma magnitud la superficie representa un valle redondo (véase figura 12.9). Dado que la superficie resulta ser del tipo deseado se procede a determinar el punto estacionario, el cual está dado por:

$$\mathbf{x}_0 = \frac{-\mathbf{B}^{-1}\mathbf{b}}{2} = \frac{-\begin{pmatrix} 0.562 & 0 \\ 0 & 0.562 \end{pmatrix}\begin{pmatrix} -0.83 \\ 1.50 \end{pmatrix}}{2} = \begin{pmatrix} 0.23 \\ -0.42 \end{pmatrix}$$

y como se esperaba, es prácticamente idéntico al obtenido de manera gráfica. Además, este punto se ubica dentro de la región experimental, por lo tanto es un punto óptimo.

De regreso a la dirección de pérdida mínima. En algunos problemas de optimización, al implementar el punto óptimo en el proceso sucede que, por limitaciones, éste no puede fijarse en estas condiciones de operación, por lo que surge la necesidad de encontrar el siguiente mejor punto donde sí es posible correr el proceso. Este punto se ubica en la dirección de mínima pérdida, dada por la variable canónica que tiene el menor coeficiente λ en valor absoluto. Esto se puede resolver en forma matemática o gráfica. En el último caso, se explora con detalle la superficie de respuesta y, con base en esto y en el conocimiento del proceso mismo, se propone un punto alternativo donde se pronostique que el proceso tendrá un rendimiento satisfactorio y será factible.

Análisis de cordillera

Muchas veces, el punto estacionario no es del tipo que se requiere (véase la figura 12.11), y en esos casos la opción es encontrar el "mejor punto posible" dentro de la región experimental. Este punto se ubica sobre la cordillera óptima a partir del centro

del diseño, y es aquel que predice la mejor respuesta sobre la región. Esta búsqueda se hace precisamente con el llamado *análisis de cordillera*, que consiste en calcular el máximo o mínimo de la superficie de respuesta sobre esferas concéntricas al centro del diseño, empezando por una esfera de radio casi cero y posteriormente se abre la exploración incrementando el radio de la siguiente esfera. Así se continúa hasta llegar a la primera esfera que cubre los puntos experimentales. El mejor punto posible es aquel sobre el que se predice el óptimo desempeño de la variable de respuesta. Con frecuencia, este punto se ubica en la esfera de radio más grande. En el caso de $k = 2$ factores, no son esferas sino circunferencias como en las de la figura 12.12. Note que en esta figura se van alcanzando mejores puntos y se va escalando la superficie. Asimismo, en cada paso se corrige el rumbo debido a la curvatura de la superficie.

Análisis de cordillera
Técnica que se emplea para determinar el mejor punto dentro de la región experimental.

Ahora, brevemente veamos en forma matemática el análisis de cordillera. Consideremos el modelo ajustado de segundo orden escrito en su forma matricial

$$\hat{Y} = \hat{\beta}_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}$$

donde **b** y **B** se construyen como en (12.4). Sea la esfera centrada en el origen con radio R_i, cuyos puntos sobre ella cumplen la restricción:

$$\sum_{i=1}^{k} x_i^2 = \mathbf{x}'\mathbf{x} = R_i \qquad \textbf{(12.9)}$$

El problema del análisis de cordillera es encontrar el punto sobre la esfera, donde la respuesta predicha por el modelo es máxima (o mínima). Para ello se plantea la función objetivo dada por

$$F = \hat{\beta}_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x} - \lambda(\mathbf{x}'\mathbf{x} - R_i)$$

donde λ es multiplicador de Lagrange. Derivando esta última relación con respecto al vector **x** e igualando a cero, se obtiene

$$\frac{\partial F}{\partial \mathbf{x}} = \mathbf{b} + 2\mathbf{B} - 2\lambda\mathbf{x} = 0$$

y de aquí se llega al sistema de ecuaciones

$$(\mathbf{B} - \lambda\mathbf{I})\mathbf{x} = \frac{-\mathbf{b}}{2} \qquad \textbf{(12.10)}$$

El punto $(x_1, x_2, ..., x_k)$ óptimo sobre una esfera particular se encuentra al sustituir un valor para λ, que no sea un valor propio de la matriz **B** en esta última relación, y se resuelve el sistema de ecuaciones resultante.

En general es mejor recurrir a un software para hacer el análisis de cordillera. Por ejemplo, en *Statgraphics*, una vez que el modelo ha sido ajustado y depurado, se elige la opción de *Optimization* y ahí, haciendo clic con el botón derecho del *mouse*, se puede indicar si se quiere minimizar, maximizar o elegir un valor objetivo. También se le especifica el rango de valores de los factores en donde se quiere la optimización.

A continuación presentamos un ejemplo en donde se aplica tanto la técnica de escalamiento ascendente como el análisis canónico.

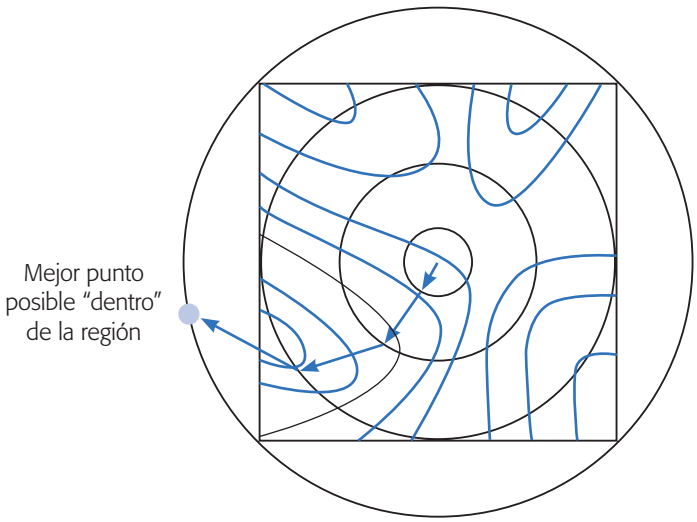


Figura 12.12 Representación gráfica del análisis de cordillera.

Ejemplo 12.4

En Hao *et al*. (2006) se aplica la MSR para evaluar el efecto de los componentes del medio de cultivo en la producción de *cellulase* (celulosa hidrolizada) mediante una mutación del hongo *Trichoderma reesei* WX-112, así como para buscar la optimización del medio de cultivo y lograr altos rendimientos en la producción de *cellulase*. Para determinar los componentes que podrían ser importantes en la producción se utilizó un diseño factorial fraccionado 2^{6-2} con cuatro puntos al centro, teniendo los siguientes factores y niveles:

Componente	Símbolo		Niveles (en g/L)		
	Real	Código	−1	0	1
Salvado de trigo	Z_1	X_1	20	30	40
Avicel	Z_2	X_2	15	25	35
Harina de soja (soya)	Z_3	X_3	10	20	30
KH_2PO_4	Z_4	X_4	2	4	6
Extracto de levadura	Z_5	X_5	5	10	15
Harina de maíz	Z_6	X_6	2	5	8

En la tabla 12.6 se muestran los tratamientos corridos y los resultados obtenidos. La variable de respuesta fue la actividad enzimática (*Filter paper activity*, FPA) medida en unidades internacionales. Al ajustar un modelo de primer orden a estos datos se obtiene el ANOVA de la tabla 12.7. Se aprecia que los únicos componentes que tuvieron una influencia significativa sobre Y fueron x_2 y x_3. Los coeficientes de determinación (R^2) son satisfactoriamente altos. Además, si consideramos sólo estos factores no hay evidencia de curvatura.

El modelo ajustado con sólo términos significativos está dado por:

$$Y = 7.515 + 0.82x_2 + 0.37x_3 \qquad \textbf{(12.11)}$$

Tabla 12.6. Diseño 2^{6-2} y resultados para el ejemplo 12.4.

Tratamiento	X_1	X_2	X_3	X_4	X_5	X_6	Y
1	–1	–1	–1	–1	–1	–1	6.0
2	1	–1	–1	–1	1	–1	6.1
3	–1	1	–1	–1	1	1	7.8
4	1	1	–1	–1	–1	1	8.4
5	–1	–1	1	–1	1	1	7.2
6	1	–1	1	–1	–1	1	7.4
7	–1	1	1	–1	–1	–1	9.0
8	1	1	1	–1	1	–1	7.9
9	–1	–1	–1	1	–1	1	6.5
10	1	–1	–1	1	1	1	6.9
11	–1	1	–1	1	1	–1	8.6
12	1	1	–1	1	–1	–1	7.5
13	–1	–1	1	1	1	–1	7.8
14	1	–1	1	1	–1	–1	6.3
15	–1	1	1	1	–1	1	9.0
16	1	1	1	1	1	1	9.1
17	0	0	0	0	0	0	7.3
18	0	0	0	0	0	0	7.1
19	0	0	0	0	0	0	7.2
20	0	0	0	0	0	0	7.2

La superficie de respuesta en regiones de contornos para este modelo se muestra en la figura 12.13*a*. De acuerdo con lo que se ha explicado, resulta apropiado determinar la trayectoria de máximo ascenso para optimizar el medio de cultivo. Para ello, con respecto a los factores que no influyeron de forma significativa, se decidió dejar fijos a X_1, X_4 y X_6 en su nivel intermedio, y como X_5 resultaba ser un componente caro, se decidió excluirlo del medio de cultivo en los siguientes experimentos. La trayectoria de máximo crecimiento para Y está definida por el modelo (12.11), el cual indica que por cada 0.82 unidades de incremento de X_2 se deberán aumentar 0.37 unidades a X_3, esto en unidades codificadas.

Tabla 12.7. Análisis de varianza para diseño de tabla 12.6.

FV	*SC*	*GL*	*CM*	F_0	Valor-*p*
x_1	0.3306	1	0.3306	1.57	0.2326
x_2	10.726	1	10.726	50.86	0.0000
x_3	2.176	1	2.176	10.32	0.0068
x_4	0.2256	1	0.2256	1.07	0.3198
x_5	0.1056	1	0.1056	0.50	0.4916
x_6	0.6006	1	0.6006	2.85	0.1153
Error	2.742	13	0.2109		
Total	16.906	19			
$R^2 = 83.8\%$, $R^2_{(aj)} = 76.3\%$					

Para determinar la longitud de paso en la trayectoria se utilizó a X_2, puesto que fue el factor más importante (ver F_0 en la tabla 12.7). Asimismo, se decidió que Z_2 se incrementará en 2.5 g/L cada vez (en unidades reales). Nótese que es una longitud de paso conservador, porque inicialmente ese factor tenía niveles dados por 25 ± 10. Para determinar el incremento en unidades codificadas que implica un incremento en Z_2 de 2.5, se aplica la siguiente fórmula:

$$\nabla X_2 = \frac{\nabla Z_2}{(Z_H - Z_L)/2} = \frac{2.5}{(35-15)/2} = \frac{2.5}{10} = 0.25$$

donde Z_H y Z_L son los niveles alto (+1) y bajo (–1) en unidades originales. Así, en unidades codificadas X_2 se incrementará 0.25 cada vez. El incremento del otro factor se obtiene con los coeficientes del modelo (12.11) de la siguiente manera:

$$\nabla X_3 = \frac{0.37}{0.82}\nabla X_2 = 0.45 \times 0.25 = 0.11$$

que en unidades reales equivale a:

$$\nabla Z_3 = \nabla X_3 (Z_H - Z_L)/2 = 0.11(30-10)/2 = 1.1$$

Con esta información se obtiene la trayectoria (los tratamientos a correr). En la tabla 12.8 se muestra esta trayectoria junto con los resultados obtenidos experimentalmente. Ahí se aprecia que después de la corrida seis se empieza a dar un cambio en la trayectoria, al empezar a descender Y.

Por lo anterior, los niveles usados en la corrida seis de la tabla 12.8 serán el centro de la nueva región experimental (ver figura 12.13), la cual se explora con un diseño 2^2 con cinco puntos al centro. Los niveles (–1,+1) en unidades reales para los dos factores son determinados de acuerdo con Z_2: 37.5 ± 5.5 y Z_3: 25.5 ± 5. El diseño y los resultados se muestran en la parte indicada en la tabla 12.9. Al ajustar a estos datos un modelo de primer orden con interacción, resulta que el modelo no describe de manera satisfactoria los resultados experimentales. La razón de ello es que hay una fuerte curvatura, como queda claro con la prueba de falta de ajuste del ANOVA

Tabla 12.8 Trayectoria de máximo ascenso y resultados obtenidos.

Corrida	X_2	X_3	Z_2	Z_3	Y
Long. de paso	0.25	0.11	2.5	1.1	
1 (Origen)	0.0	0.0	25.0	20.0	7.2
2	0.25	0.11	27.5	21.1	7.8
3	0.5	0.22	30.0	22.2	8.6
4	0.75	0.33	32.5	23.3	9.2
5	1.0	0.44	35.0	24.4	10.2
6	1.25	0.55	37.5	25.5	10.6
7	1.5	0.66	40.0	26.6	10.3
8	1.75	0.77	42.5	27.7	9.4
9	2.0	0.88	45.0	28.8	8.6

Tabla 12.9 Diseño y resultados para la segunda región experimental, ejemplo 12.4.

Tratamientos	X_2	X_3	Z_2	Z_3	Y
Diseño 2^2 con puntos centrales	–1	–1	32	20.5	7.9
	1	–1	43	20.5	7.1
	–1	1	32	30.5	6.4
	1	1	43	30.5	6.5
	0	0	37.5	25.5	10.6
	0	0	37.5	25.5	10.6
	0	0	37.5	25.5	10.4
	0	0	37.5	25.5	10.5
	0	0	37.5	25.5	10.2
+ puntos axiales	–1.41	0	29.7	25.5	9.1
	1.41	0	45.2	25.5	6.1
	0	–1.41	37.5	18.5	8
	0	1.41	37.5	32.3	6.3

de la parte superior de la tabla 12.10. Por ello, es necesario extender el diseño original a un diseño de segundo orden, ya que así se completa un diseño rotable de composición central (ver sección siguiente). En la tabla 12.9 se agregaron los llamados *puntos axiales* y también se muestran los resultados obtenidos. Para encontrar los niveles reales a los que corresponde el código $X_i = 1.41$ (o –1.41), se aplicó la ecuación (12.7):

$$Z_i = \frac{x_i(Z_H - Z_L) + (Z_H + Z_L)}{2} \qquad \textbf{(12.7)}$$

Al ajustar un modelo de segundo orden a todos los datos de la tabla 12.9 se obtiene el ANOVA de la parte baja de la tabla 12.10, en donde se aprecia que este modelo describe adecuadamente la superficie. Eliminando el término X_2X_3 que es no significativo, se obtiene que el modelo ajustado está dado por:

$$Y = 10.46 - 0.62X_2 - 0.56X_3 - 1.53X_2^2 - 1.755X_3^2$$

La superficie de respuesta en contornos para este modelo se muestra en la figura 12.13*b*, donde se aprecia claramente que se tiene un máximo cerca de la zona central de la región experimental. El óptimo está en el punto (X_2, X_3) = (–0.202, –0.161), en donde se predice un nivel de Y igual a 10.6. Para encontrar a qué condiciones reales corresponde este punto se aplica la ecuación (12.7) y se obtiene que (Z_2, Z_3) = (36.4, 24.7). Para verificar lo adecuado de este punto se realizaron tres corridas confirmatorias en este punto y los valores de Y fueron 10.5, 10.7 y 10.6. De esta manera, si se considera que en el centro de la región experimental inicial los valores promedio de Y fueron 7.2 (ver tabla 12.6), y en el óptimo se alcanzan valores de 10.6, entonces se logró incrementar 47% el rendimiento con la aplicación de la metodología, además se eliminó un componente caro (el X_5). En la figura 12.13 se aprecia gráficamente todo el proceso.

Tabla 12.10 Análisis de varianza para el modelo de primer y segundo orden.

ANOVA modelo de primer orden

FV	SC	GL	CM	F_0	Valor-*p*
X_2	0.1225	1	0.1225	4.37	0.1046
X_3	1.1025	1	1.103	39.37	0.0033
X_2X_3	0.2025	1	0.2025	7.23	0.0547
Falta de ajuste	26.990	1	26.989	963.9	0.0000
Error puro	0.112	4	0.028		
Total	28.529	8			
$R^2 = 5.0\%$ $R^2_{(aj)} = 0.0\%$					

ANOVA modelo de segundo orden

FV	SC	GL	CM	F_0	Valor-*p*
X_2	3.054	1	3.054	10.62	0.0139
X_3	2.536	1	2.536	8.82	0.0208
X_2X_2	16.284	1	16.284	56.65	0.0001
X_2X_3	0.2025	1	0.2025	0.70	0.4290
X_3X_3	21.426	1	21.426	74.53	0.0001
Error	2.012	7	0.2875		
Total	41.211	12			
$R^2 = 95.1\%$ $R^2_{(aj)} = 91.6\%$					

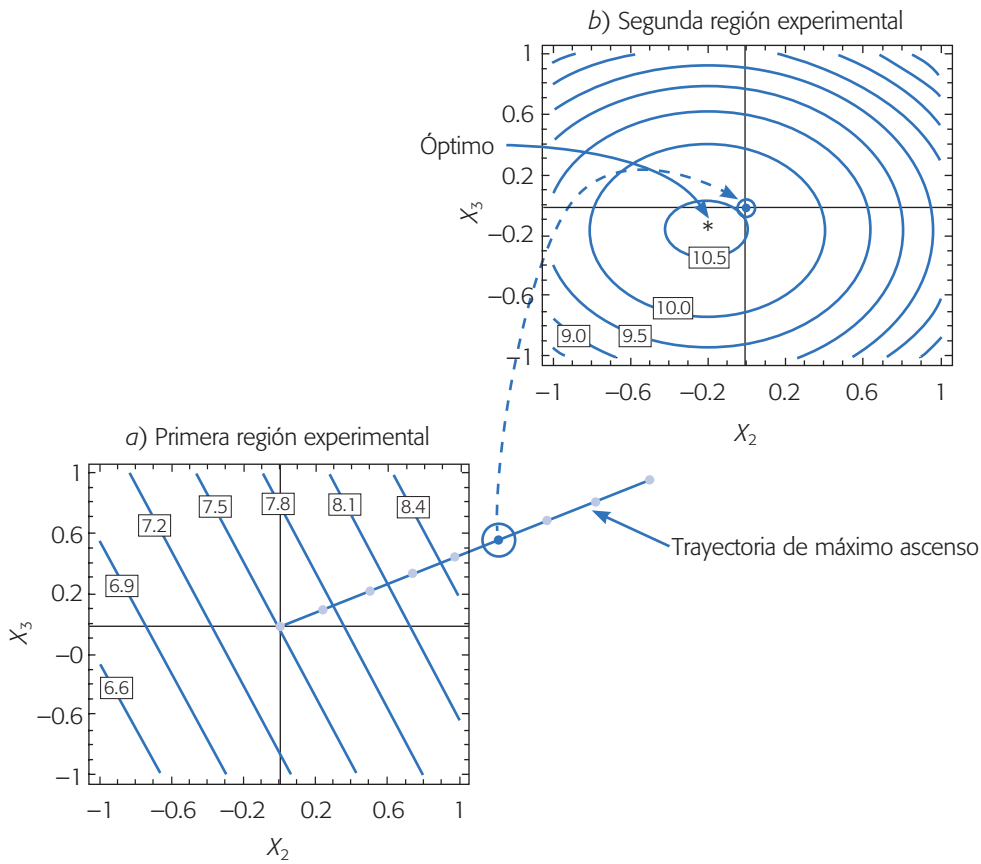


Figura 12.13 Superficies de respuesta para el problema de optimización del ejemplo 12.4.

Diseños de superficie de respuesta

Diseños de superficie de respuesta
Diseños experimentales utilizados cuando el objetivo es ajustar un modelo para describir una superficie de respuesta.

Los *diseños de superficie de respuesta* se clasifican con base en el grado del modelo que se pretende utilizar. Estos diseños proporcionan los tratamientos a correr para generar datos que permitan ajustar un modelo que describa una variable de respuesta en una región experimental. Algunas propiedades deseables en los diseños para la MSR son:

1. Que genere una distribución satisfactoria de los puntos experimentales sobre la región experimental. Los diseños más utilizados son puntos distribuidos de manera uniforme sobre la región experimental, o cuando menos tienen alguna simetría con respecto al centro de ésta.
2. El diseño debe requerir un número mínimo de corridas experimentales, ya que en cada prueba realizada se gastan recursos que siempre son escasos.
3. El diseño debe permitir que otros diseños de orden mayor se construyan a partir de él. Esto permite que, cuando el comportamiento de la respuesta resulta ser más complicado de lo que se pensaba (por ejemplo, se detecta curvatura), se agregan puntos adicionales al diseño para tratar de explicar ese comportamiento.
4. El experimento debe permitir la detección de la falta de ajuste, para lo cual se requieren repeticiones al menos en el centro del diseño.
5. El diseño debe proporcionar un estimador puro de la varianza del error, lo cual se logra con repeticiones al menos en el punto central.

Otras dos propiedades deseables en los diseños para superficie de respuesta son la *ortogonalidad* y la *rotabilidad*. Estas propiedades aumentan la eficiencia de los diseños que las poseen, en el sentido de que facilitan la interpretación de los parámetros estimados en el modelo y de la superficie de respuesta.

Ortogonalidad y rotabilidad

Diseño ortogonal
Cuando las columnas de la matriz de diseño son independientes entre sí, hace que los coeficientes del modelo ajustado no estén correlacionados.

Se considera que un *diseño* es *ortogonal* cuando los coeficientes estimados en el modelo ajustado no están correlacionados entre sí, lo cual hace que el efecto de cada término, representado por el parámetro correspondiente, se estime de manera más precisa. Un *experimento* es *ortogonal* si en la matriz de diseño todos los vectores columna son independientes entre sí. Es fácil verificar que en un diseño factorial completo 2^k las columnas de su matriz de diseño son independientes: multiplique dos columnas cualesquiera, término a término usando la notación –1 y +1, y el resultado es cero.

Diseño rotable
Es aquel en el que la varianza de la respuesta estimada sobre un punto depende de la distancia de éste al centro del diseño y no de la dirección en la que se encuentra.

Un *diseño* se llama *rotable* si la varianza de $\hat{Y}(x)$ sólo depende de la distancia del punto $\boldsymbol{x}$ al centro del diseño y no de la dirección en la que se encuentra. Es decir, si pensamos en la variable var$[\hat{Y}(x)]$ como otra respuesta, su gráfica de contornos tiene la forma de círculos concéntricos alrededor del centro del diseño. La rotabilidad del diseño asegura que la calidad de la predicción, medida por var$[\hat{Y}(x)]$, sea

invariante a cualquier rotación del diseño alrededor del centro del mismo; de aquí se deriva el nombre de esta propiedad.

La importancia práctica de la rotabilidad en el problema de encontrar un punto óptimo es porque asegura que, con independencia de la dirección en que se explore a partir del centro del diseño, la calidad de predicción se comporta de igual manera.

Relación modelo-diseño

Existe una relación directa entre el tipo de modelo que se pretende ajustar y el tipo de diseño que se debe correr. No se debe exigir a un diseño experimental más información de la que puede dar. Por ejemplo, si se corre un factorial completo 2^k sólo se podrán estimar e incluir en el modelo los efectos principales e interacciones dobles; asimismo, no es posible estimar términos cuadráticos puros (como x_i^2). Si al diseño factorial se le agregan repeticiones al centro (2^k + centro), en el modelo se puede incluir sólo uno de los términos cuadráticos puros, cualquiera de ellos, ya que son alias. Las repeticiones al centro no son suficientes para investigar cuál o cuáles de los términos cuadráticos están activos, pero sí permiten detectar la presencia de curvatura. Si el diseño se aumenta con puntos estrella o axiales es posible estudiar de manera separada los efectos cuadráticos puros e incluirlos a todos en el modelo ajustado, si fuera necesario (véase figura 12.14).

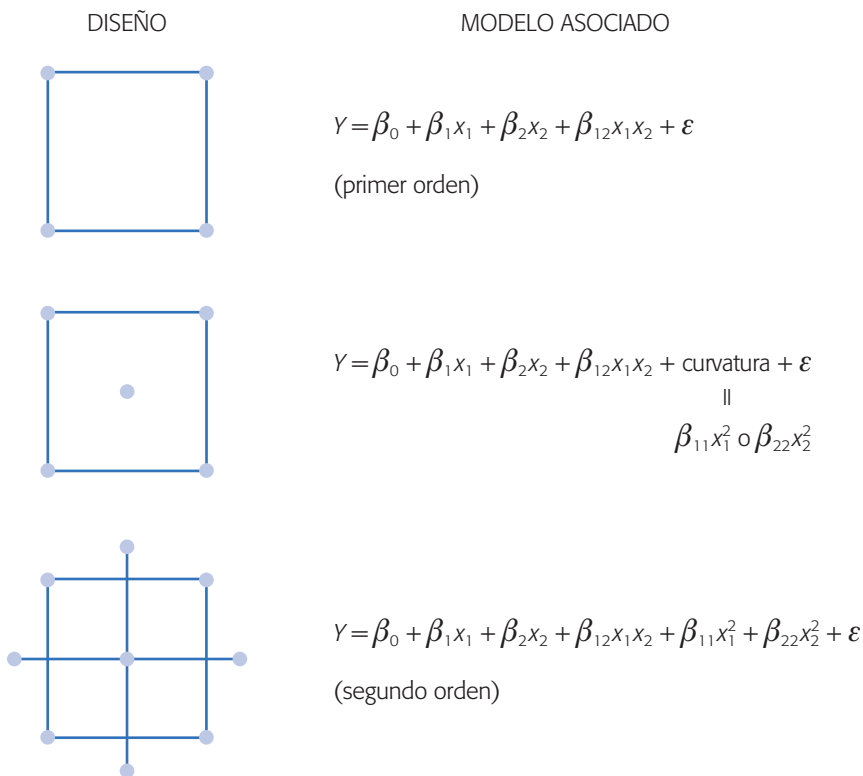

DISEÑO
MODELO ASOCIADO
$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$
(primer orden)
$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \text{curvatura} + \varepsilon$
$\beta_{11} x_1^2$ o $\beta_{22} x_2^2$
$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$
(segundo orden)

Figura 12.14 Relación modelo-diseño.

Diseños de primer orden

Suponga que se desea utilizar el modelo de primer orden dado por la ecuación (12.3) para estudiar el comportamiento de cierta característica de calidad, que se supone depende de k factores de proceso. En principio, al proponer un *diseño de primer orden* se supone que sólo son importantes los efectos principales. Estrictamente hablando, para estimar los $k + 1$ parámetros del modelo de primer orden se requiere un mínimo de $k + 1$ puntos experimentales.

Diseño de primer orden
Diseño para ajustar modelos en los que sólo son importantes los efectos principales y no existen efectos de interacción.

Un criterio de selección del diseño de primer orden es que la varianza de la respuesta predicha (var$[\hat{Y}(x)]$) en el punto $x' = (x_1, x_2,..., x_k)$ sea mínima. Este criterio es importante porque cuando se busca determinar la dirección óptima de movimiento a partir de los predichos por el modelo, éstos tienen mayor precisión, lo cual se traduce en mayor certeza de la dirección seleccionada. Los diseños que satisfacen este criterio son los que tienen la propiedad de ortogonalidad. Entre los más utilizados están los siguientes:

1. Diseños factoriales 2^k
2. Diseños factoriales fraccionados 2^{k-p}
3. Diseño de Plackett-Burman
4. Diseño simplex

Todos estos diseños, excepto el diseño simplex, emplean dos niveles en cada factor, lo cual tiene que ver con el hecho de que sólo interesa detectar el efecto principal de cada factor. Sin embargo, una vez superada la etapa de cribado es importante aumentar estos arreglos con repeticiones al centro a fin de detectar la presencia de curvatura o falta de ajuste del modelo; las repeticiones al centro también proporcionan más grados de libertad para el error aleatorio. Lo diseños 1 a 3 se explicaron en los capítulos previos.

El *diseño simplex* para k factores se representa por medio de una figura de forma regular dibujada en un espacio de dimensión $k - 1$, y se caracteriza por el hecho de que el ángulo θ formado por cualquier par de vértices con el origen, es tal que $\cos(\theta) = -1/k$. Así, para $k = 2$ factores, los tratamientos del diseño simplex corresponden a los vértices de un triángulo equilátero (véase figura 12.15); para $k = 3$ son los vértices de un tetraedro (figura 12.15). Las matrices de diseño en unidades codificadas para estos dos casos también se muestran en la figura. Observe que algunos factores se prueban en dos niveles y otros en tres.

Diseño simplex
Arreglo que se utiliza cuando se tienen muchos factores como en la etapa de cribado, o bien, cuando se quiere minimizar el costo de la experimentación.

En el capítulo 15 se emplea una variante especial de diseños simples para estudiar experimentos con mezclas.

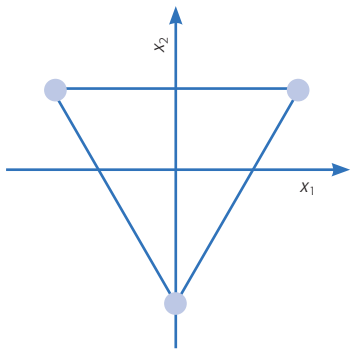
Diseños de segundo orden

Se llaman *diseños de segundo orden* aquellos que permiten ajustar un modelo de segundo orden para así estudiar, además de los efectos lineales y de interacción, los efectos cuadráticos o de curvatura pura. Por consiguiente, estos diseños se emplean cuando se quiere explorar una región que se espera sea más compleja o cuando se cree que el punto óptimo ya se encuentra dentro de la región experimental. El modelo de segundo orden está dado por:

Diseños de segundo orden
Permiten estudiar efectos lineales, de interacción y efectos cuadráticos o de curvatura pura.

X_1	X_2
$\sqrt{3/2}$	$1/\sqrt{2}$
$-\sqrt{3/2}$	$1/\sqrt{2}$
0	$-2/\sqrt{2}$

X_1	X_2	X_3
$-\sqrt{2}$	$-\sqrt{3/2}$	$-1/\sqrt{3}$
$\sqrt{2}$	$-\sqrt{3/2}$	$-1/\sqrt{3}$
0	$\sqrt{6}$	$-1/\sqrt{3}$
0	0	$\sqrt{3}$



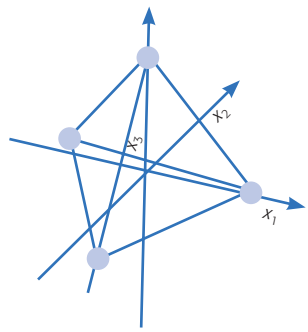


Figura 12.15 Diseños simplex, para $k = 2$ y 3 factores.

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1<j=1}^{k}\sum^{k} \beta_{ij} x_i x_j + \varepsilon$$

tiene $p = (k + 1)(k + 2)/2$ términos, por lo tanto se requiere al menos esa cantidad de puntos de diseño. El diseño debe tener al menos tres niveles en cada factor para estimar la curvatura de la superficie en la dirección de cada factor. Es deseable que estos diseños sean ortogonales, pero a veces no es fácil que cumplan esta propiedad y se admite alguna dependencia entre las columnas de los contrastes. Los más utilizados tienen la propiedad de ser rotables.

A continuación se presentan las matrices de diseño, la geometría y propiedades de los diseños de segundo orden más recomendados como son: el diseño de *Box-Behnken* y el *diseño central compuesto* o de composición central. Debemos decir que los diseños 3^k presentados en el capítulo 7 también sirven para ajustar un modelo de segundo orden, pero no son los más recomendados porque requieren más corridas experimentales.

Diseños de Box-Behnken
Diseños de segundo orden para tres o más factores. No incluye como tratamientos a los vértices de la región experimental.

Diseño de Box-Behnken. Este *diseño* se aplica cuando se tienen tres o más factores, y suelen ser eficientes en cuanto al número de corridas. Es un diseño rotable o casi rotable que se distingue porque no incluye como tratamientos a los vértices de la región experimental. En la tabla 12.11 se muestran los 15 tratamientos del diseño Box-Behnken para tres factores.

Su representación geométrica se muestra en la figura 12.16. Note que los puntos de este diseño se ubican en medio de las aristas del cubo centrado en el origen y, como se había señalado, no incluye los tratamientos de los vértices, como el (1, 1, 1) y el (–1, –1, –1) que en algunas situaciones experimentales resultan extremosos y no se pueden correr. Esto hace que los tratamientos sean menos extremos que los del

Tabla 12.11 Matriz del diseño de Box-Behnken para tres factores.

Tratamiento	x_1	x_2	x_3	Tratamiento	x_1	x_2	x_3
1	–1	–1	0	9	–1	0	–1
2	1	–1	0	10	1	0	–1
3	–1	1	0	11	–1	0	1
4	1	1	0	12	1	0	1
5	0	–1	–1	13	0	0	0
6	0	1	–1	14	0	0	0
7	0	–1	1	15	0	0	0
8	0	1	1				

factorial completo 3^2. En cualquier tratamiento del diseño de Box-Behnken al menos uno de los factores se fija en la mitad de su rango de prueba. Lo anterior también ocurre con este diseño para más factores.

Diseño de composición central. El *diseño de composición central* (DCC) es el más utilizado en la etapa de búsqueda de segundo orden debido a su gran flexibilidad: se puede construir a partir de un diseño factorial completo 2^k o fraccionado 2^{k-p} agregando puntos sobre los ejes y al centro (véase ejemplos 12.3 y 12.4), además de otras propiedades deseables. Este diseño se compone de tres tipos de puntos:

1. Una réplica de un diseño factorial en dos niveles, completo o fraccionado. A esta parte del DCC se le llama *porción factorial*.
2. n_0 puntos o repeticiones al centro del diseño, con $n_0 \geq 1$.
3. Dos puntos sobre cada eje a una distancia α del origen. Estos puntos se llaman *porción axial*. La manera en que se ensamblan y el DCC resultante se muestran en figura 12.17, para los casos de dos y tres factores.

Diseño de composición central
Se emplea en la etapa de búsqueda de segundo orden. Se puede construir a partir de un diseño factorial 2^k agregando puntos sobre los ejes y al centro del diseño.

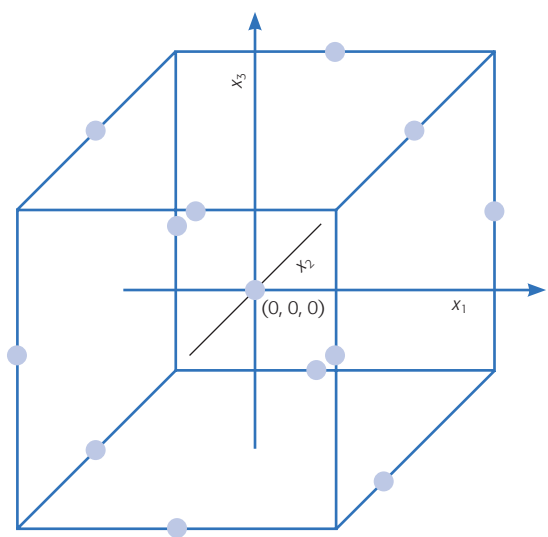


Figura 12.16 Representación del diseño de Box-Behnken para tres factores.

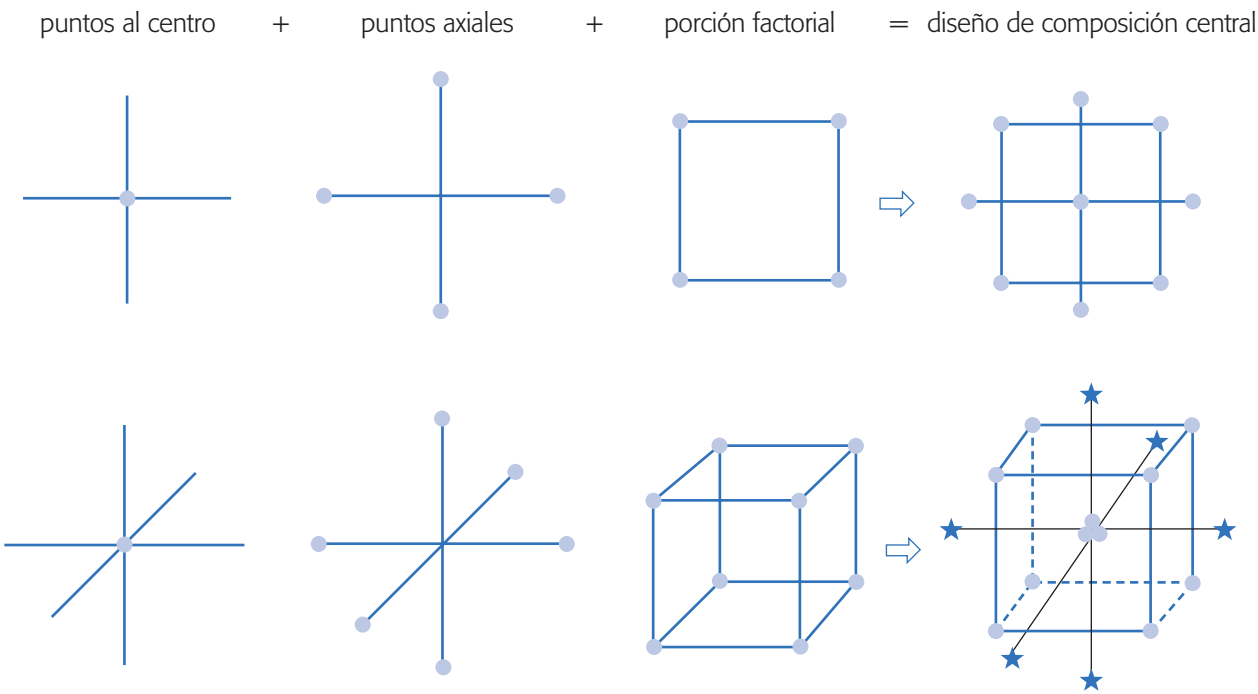


Figura 12.17 Representación y construcción de los DCC para $k = 2$ y 3 factores.

Con k factores el número total de corridas experimentales es $N = 2^k + 2k + n_0$. El número de corridas al centro y la distancia de los puntos axiales (α) deben escogerse de manera adecuada, dependiendo de las propiedades que se quieren en el DCC. Si se quiere que el diseño sea rotable se toma α igual a:

$$\alpha = (F)^{1/4}$$

donde F es el número de puntos en la parte factorial. Si se desea que el DCC sea *ortogonal* se toma α igual a:

$$\alpha = \left(\frac{(F \times N)^{1/2} - F}{2} \right)^{1/2}$$

donde N es el total de corridas experimentales. Si se desea que el DCC sea *ortogonal* y *rotable*, se deben cumplir las relaciones,

$$\alpha = (F)^{1/4} \qquad \text{y} \qquad n_0 \approx 4\sqrt{F} + 4 - F$$

Las repeticiones al centro también se pueden escoger de manera que el diseño sea *rotable* y *de precisión uniforme*. Un diseño es de precisión uniforme si var$[\hat{Y}(x)]$ en el centro del diseño (radio $r = 0$), es igual a la varianza en la esfera de radio $r = 1$. Esta propiedad proporciona aproximadamente un valor constante de la var$[\hat{Y}(x)]$ dentro de la esfera unitaria (de radio uno). La elección para rotabilidad y precisión uniforme, o cercanamente uniforme es:

$$\alpha = (F)^{1/4} \qquad \text{y} \qquad n_0 = \lambda_4 (\sqrt{F} + 2) - F - 2k$$

donde el valor de la constante λ_4 dado el número de factores k, se obtiene de la siguiente tabla:

1	2	3	4	5	6	7	8
λ_4	0.7844	0.8385	0.8704	0.8918	0.9070	0.9184	0.9274

Diseño de composición central con centros en las caras. En este diseño los puntos axiales se localizan en el centro de las caras, es decir, $\alpha = 1$. Esta variante resulta particularmente útil cuando ya se corrió el diseño 2^k y se quiere completar un DCC, pero se complica correr los tratamientos con niveles más allá de los valores –1 y +1 debido a restricciones del proceso.

Diseño central compuesto pequeño. Draper y Lin (1990) proponen un diseño central compuesto cuya parte factorial se construye a partir de un factorial fraccionado que tiene resolución menor que V, las corridas axiales y al centro se determinan como el DCC. En cuanto al número de corridas, este diseño es sin duda una de las opciones más económicas que se pueden tener para ajustar el modelo de segundo orden (véase tabla 12.12). Los tratamientos se pueden generar con un software, por ejemplo *Statgraphics* incluye este diseño como opción. La desventaja de este diseño es que algunos efectos de interés pueden ser alias.

¿Cuál diseño de segundo orden utilizar?

Existen varios criterios que se pueden tomar en cuenta para seleccionar un diseño de segundo orden, entre los que destacan:

1. Que tenga un número mínimo de tratamientos y que permita estimar todos los parámetros del modelo de segundo orden completo (tabla 12.12).
2. Que sea flexible, en el sentido de que se pueda construir a partir de diseños de primer orden.
3. Que tenga buenas propiedades como *ortogonalidad*, *rotabilidad* y/o *precisión uniforme*, las cuales tienen que ver con la calidad de estimación con el modelo.

Tabla 12.12 Número de puntos en los diseños de segundo orden.

Número de factores	Número de corridas				
	Número de parámetros	Factorial completo 3^k	Central compuesto	Box-Behnken	Draper y Lin
2	6	9	13-16	—	—
3	10	27	17-23	15	12
4	15	81	27-36	27	18
5	21	243	29-36	46	24
6	28	729	47-59	54	30

En cuanto a economía, el diseño 3^k no compite con los de Box-Behnken y DCC, excepto cuando se tienen dos factores. El diseño central compuesto es el más flexible y se puede construir de manera que cumpla alguna o varias propiedades deseables. El diseño de Box-Behnken es una buena alternativa porque tiene un número competitivo de tratamientos, además de que por su construcción los puntos de prueba no son extremosos, en el sentido de que cualquiera de ellos siempre tiene, al menos uno de los factores en su nivel intermedio.

La flexibilidad del diseño central compuesto es en el sentido de que se puede construir a partir de aumentar diseños más simples. Esta propiedad es de suma importancia porque permite optimizar el gasto de los recursos experimentales, es decir, si ya se corrió un factorial completo y con punto central, y se detectó la presencia de curvatura, ahora se corren los puntos axiales para completar el diseño. No hay necesidad de comenzar de cero, sino que se aprovechan las corridas que ya se han hecho. Aquí es clave, al hacer las corridas adicionales, poder reproducir fielmente las condiciones ambientales que se tenían antes, y en cualquier caso se debe vigilar que no aparezca algún efecto de bloque imprevisto. Por lo anterior, se prefiere el diseño central compuesto, seguido del Box-Behnken y en última instancia el factorial 3^k.

Los diseños de Draper y Lin son diseños centrales compuestos con un número mínimo de corridas experimentales. En la tabla 12.12 se reporta el número mínimo de corridas que se tienen al considerar dos repeticiones al centro. Estos diseños son útiles en una situación en la que desde el primer diseño se quieren correr los puntos necesarios para ajustar el modelo de segundo orden de manera económica, debido a lo caro de las corridas experimentales, y se está dispuesto a asumir el riesgo de que algunos efectos potencialmente importantes sean alias.

Uso de software estadístico

Los diseños de superficie de respuesta se construyen en *Statgraphics* con la secuencia *Special* → *Experimental design* → *Create design* → *Response surface*. En versiones más recientes la secuencia inicia con *DOE*. Se declaran los factores y las respuestas, se elige un diseño del menú típicamente un central compuesto (2^k + *Star)* o uno de Box-Behnken. Si el problema es agregar los puntos axiales a un diseño factorial con repeticiones al centro se usa la secuencia: *Special* → *Experimental design* → *Augment* → *Add star points*.

En *Statgraphics*, una vez seleccionado un modelo de regresión de segundo orden que describe adecuadamente el comportamiento de la respuesta (Y), la optimización de ésta se lleva a cabo en la opción tabular de *Optimization.* Una vez ahí, con las opciones de panel (botón derecho del ratón) se elige el tipo de optimización deseada: *máximo, mínimo* o *un valor objetivo*. Si se quiere un valor objetivo se declara su valor en el campo correspondiente. Se indica el rango de búsqueda para cada factor, típicamente el rango utilizado en la región experimental. Entonces, se corre el procedimiento y éste reporta las coordenadas del punto óptimo.

En caso de no existir un óptimo natural o que se pida un valor objetivo, el software reporta el mejor punto posible sobre la cordillera óptima dentro de la región experimental. En general, esta solución es el punto más cercano al centro de la región experimental que predice el valor deseado de la respuesta. Sin embargo, debe

tenerse presente que ésta es sólo una de muchas soluciones posibles. Usando gráficas de contornos (en *Response plots*) es posible encontrar soluciones alternativas que resultan más económicas de operar en el proceso.

En el caso de los diseños de primer orden, para aplicar escalamiento ascendente, éstos se eligen en la opción *Screening*. Una vez ajustado y depurado el modelo, en las opciones tabulares se incluye la opción de máximo ascenso.

En Minitab, la secuencia para los diseños de segundo orden es: *Stat* → *DOE* → *Response* → *Surface*. Una vez que el modelo está ajustado y depurado es posible ver la superficie de respuesta y encontrar el óptimo, seleccionando después de la secuencia anterior las opciones *Contour* → *Surface plots* → *Response optimizer*.

Preguntas y ejercicios

1. ¿Cuándo se recomienda aplicar la metodología de superficie de respuesta?
2. Utilice dos factores y explique con detalle la diferencia entre la región de operabilidad de un proceso y la región de experimentación. ¿Es posible que la segunda sea más grande que la primera?
3. ¿Cuál es la diferencia entre el tratamiento ganador y el tratamiento óptimo?
4. La metodología de superficie de respuesta implica tres aspectos clave: *diseño*, *modelo* y *técnica de optimización*. Explique brevemente el significado de cada uno.
5. En la MSR se distinguen tres etapas en la búsqueda del punto óptimo, que son: *cribado*, *búsqueda I* o *de primer orden* y *búsqueda II* o *de segundo orden.* Describa cada una de estas etapas.
6. Explique qué son los modelos: de primer orden, de segundo orden y jerárquicos.
7. ¿Cuándo un diseño es ortogonal y cuándo es rotable? También explique por qué son importantes estas propiedades en los diseños.
8. Explique cuándo es conveniente aplicar los diseños de primer orden y cuáles son los principales tipos de diseños de primer orden.
9. ¿Cuándo es conveniente aplicar los diseños de segundo orden y cuáles son los principales diseños de este tipo?
10. ¿Por qué el diseño y el modelo se deben pensar al mismo tiempo? ¿Qué quiere decir esto?
11. ¿Cuál es la ventaja del diseño de composición central con respecto a los demás diseños de segundo orden?
12. Describa una virtud que tiene el diseño de Box-Behnken.
13. ¿Por qué es recomendable trabajar el análisis en unidades codificadas en lugar de las unidades originales?
14. Una de las técnicas de optimización es el escalamiento ascendente. ¿Cuándo se recomienda aplicarlo y en qué consiste?
15. ¿En qué consiste el análisis canónico y cuándo se aplica?
16. En un análisis canónico ¿cuáles son los diferentes tipos de superficies que se pueden obtener?

17. ¿Qué procede si se busca un mínimo y la superficie tiene máximo? ¿Y si es un punto silla?

18. ¿En qué situación se necesita regresar en la dirección de mínima pérdida? ¿Cómo se determina dicha dirección?

19. ¿En qué consiste el análisis de cordillera y cuándo se aplica?

20. Considere el problema de diseñar un helicóptero de papel presentado en el ejemplo 12.1. En la etapa de cribado se corrió un experimento 2_{IV}^{8-4} para estudiar ocho factores. Cada uno de los 16 helicópteros diseñados se lanza cuatro veces y el tiempo de vuelo se mide en centésimas de segundo.
    a) Dado que lo que se pretende es encontrar las dimensiones del helicóptero que maximicen el tiempo de vuelo, ¿es adecuado el diseño que se emplea? ¿Qué se puede lograr con este diseño?
    b) Como se observa, cada corrida consistió en hacer un helicóptero de acuerdo a las dimensiones que señala cada tratamiento, y lanzarlo cuatro veces. ¿Por qué los cuatro lanzamientos no se pueden considerar cuatro réplicas del tratamiento? ¿Por qué fue necesario lanzarlo cuatro veces y registrar el promedio?
    c) Analice de nuevo los datos y determine el mejor modelo de regresión que explica el tiempo de vuelo.
    d) Utilice en el factor de mayor influencia un paso de $\Delta x_i = 0.8$ en unidades codificadas, determine cinco puntos experimentales sobre la dirección óptima a partir del centro de la región experimental.
    e) ¿Qué tendría que hacer con los puntos encontrados en el inciso anterior?

21. A continuación se muestran los datos obtenidos con un diseño factorial 2^5 con cuatro repeticiones al centro:

(1) = 1 123	*d* = 3 348	*e* = 4 093	*de* = 4 190	
a = 1 786	*ad* = 3 055	*ae* = 4 517	*ade* = 4 413	
b = 1 786	*bd* = 2 509	*be* = 4 755	*bde* = 4 264	
ab = 2 359	*abd* = 2 917	*abe* = 4 316	*abde* = 4 100	
c = 982	*cd* = 4 328	*ce* = 7 066	*cde* = 6 935	(0, 0, 0, 0, 0) = 5 134
ac = 1 458	*acd* = 3 969	*ace* = 5 871	*acde* = 6 467	(0, 0, 0, 0, 0) = 5 157
bc = 1 451	*bcd* = 2 932	*bce* = 5 477	*bcde* = 5 306	(0, 0, 0, 0, 0) = 4 653
abc = 2 180	*abcd* = 4 167	*abce* = 5 774	*abcde* = 5 960	(0, 0, 0, 0, 0) = 4 834

    a) Ajuste un modelo de primer orden y compruebe que es adecuado.
    b) ¿El modelo presenta falta de ajuste?
    c) Si el modelo de primer orden es adecuado, encuentre la dirección de máximo ascenso.
    d) Determine las coordenadas de cuatro puntos en esta dirección, expresados en unidades codificadas. Use un paso unitario de movimiento en la variable de mayor influencia.
    e) Dibuje los puntos en la dirección óptima.

22. En una empresa dedicada a anodizar artículos de aluminio (baterías de cocina), el anodizado se logra en una solución a base de ácidos (sulfúrico, cítrico, bórico) y dicromato de aluminio. Debido al poco grosor del anodizado, han aumentado las quejas por la escasa resistencia y durabilidad del producto. Para resolver este problema se decide estudiar mediante un experimento la relación del pH y la temperatura con el grosor del anodizado. Los datos se muestran en la siguiente tabla:

pH	Temperatura	Espesor
1.2	–8	9
1.8	–8	14
1.2	8	10
1.8	8	19
1.2	–8	8
1.8	–8	12
1.2	8	11
1.8	8	20
1.5	0	14
1.5	0	13

a) ¿Cómo se llama el diseño empleado?
b) ¿Qué se gana con las repeticiones al centro?
c) Encuentre el mejor modelo que describe el comportamiento del espesor.
d) Con el modelo anterior liste cinco puntos en la dirección de máximo ascenso, utilice un paso de 0.3 en el factor de mayor influencia.
f) Explique qué se hace después con los puntos encontrados.

23. Los siguientes datos fueron colectados por un ingeniero químico. La respuesta *Y* es el *tiempo de filtración*, x_1 es la *temperatura* y x_2 es la *presión*.

x_1	x_2	Y
–1	–1	54
–1	1	45
1	–1	32
1	1	47
–1.414	0	50
1.414	0	53
0	–1.414	47
0	1.414	51
0	0	41
0	0	39
0	0	44
0	0	42
0	0	40

a) Represente graficamente el diseño y diga ¿qué diseño es y qué propiedades tiene?
b) Ajuste un modelo de segundo orden y compruebe que es adecuado.
c) ¿Qué tipo de superficie describe este modelo?
d) ¿Cuál es el mejor punto posible de operación?

24. Se tiene un proceso de extrusión para producir harina instantánea de amaranto. Una de las variables que interesa minimizar es el *índice de solubilidad en agua* (ISA) y los factores que se controlan son: *temperatura* (X_1), *porcentaje de humedad* (X_2) y *velocidad de tornillo* (X_3). Los datos obtenidos para la solubilidad sobre un diseño de Box-Behnken se muestran a continuación:

x_1	x_2	x_3	ISA
−1	−1	0	15.87
1	−1	0	12.70
−1	1	0	14.80
1	1	0	13.53
−1	0	−1	15.10
1	0	−1	12.47
−1	0	1	11.37
1	0	1	10.27
0	−1	−1	15.33
0	1	−1	15.53
0	−1	1	15.17
0	1	1	14.17
0	0	0	13.85
0	0	0	13.93
0	0	0	13.77

a) Represente gráficamente el diseño empleado y señale su nombre y cuándo se recomienda aplicarlo.
b) Ajuste un modelo de segundo orden a estos datos y verifique su calidad.
c) ¿Qué tipo de superficie describe el modelo?
d) Haga un análisis de cordillera para determinar el mejor punto posible dentro de la región experimental.

25. Después de aplicar experimentación secuencial se encontró una zona de experimentación donde se sospecha que se encuentran las condiciones óptimas para la producción de un colorante natural en función de la concentración de carbono y temperatura. El diseño empleado y las producciones encontradas se muestran a continuación:

Con. Carbono	Temperatura °C	Producción
9	17	5 707
13	17	5 940
9	25	3 015
13	25	2 673
8.17	21	5 804
13.8	21	6 700
11	15.34	5 310
11	26.66	725
11	21	7 521
11	21	7 642
11	21	7 500
11	21	7 545

a) Represente gráficamente en el plano el diseño empleado, utilice condiciones reales.
b) ¿Cuál es el nombre del diseño?
c) Ajuste un modelo de segundo orden y depúrelo.
d) ¿El modelo describe adecuadamente la variación observada? Argumente su respuesta.
e) Obtenga la gráfica de la superficie de respuesta y la gráfica de contornos.
f) Encuentre las condiciones que maximizan la producción del colorante.

26. Se busca encontrar los niveles de tiempo (W_1) y temperatura (W_2) que maximizan el rendimiento de un proceso. Las condiciones actuales son $W_1 = 75$ minutos y de $W_2 =$ 130°C. Para explorar la superficie de respuesta en torno a estos valores se corre el siguiente diseño experimental.

Condiciones reales		Condiciones codificadas		
W_1	W_2	X_1	X_2	Y
70	127.5	–1	–1	54.3
80	127.5	1	–1	60.3
70	132.5	–1	1	64.6
80	132.5	1	1	68.0
75	130.0	0	0	60.3
75	130.0	0	0	64.3
75	130.0	0	0	62.3

a) ¿Cuál es el diseño empleado? ¿Es adecuado para hacer una exploración inicial?

b) Ajuste el modelo de primer orden, ¿es adecuado (curvatura, R^2)? Explique.

c) Anote la ecuación del modelo con el que se encuentra la trayectoria de máximo ascenso.

d) Con base en la trayectoria de máximo crecimiento, proponga en qué niveles de W_1 y W_2 recomendaría experimentar.

Siguiendo la trayectoria de máximo crecimiento, el último punto con el que se obtuvo una respuesta alta fue $W_1 = 90$ y $W_2 = 145$, por lo que en torno a éste se corrió el siguiente diseño:

Condiciones reales		Condiciones codificadas		
W_1	W_2	X_1	X_2	Y
80	140	–1	–1	78.8
100	140	1	–1	84.5
80	150	–1	1	91.2
100	150	1	1	77.4
90	145	0	0	89.7
90	145	0	0	86.8

e) Represente en el mismo plano, en condiciones reales, a las dos zonas de experimentación (los dos diseños) que se han explorado hasta ahora.

f) A los datos del segundo diseño ajústeles un modelo de primer orden y vea si éste describe adecuadamente la superficie de respuesta.

Dado que en el inciso anterior parece que el modelo de primer orden no describe la superficie, se decide agregar seis tratamientos para completar un diseño de composición central y de esta manera explorar ampliamente la región experimental para entender lo que pasa. Las corridas agregadas y sus resultados se muestran a continuación:

Condiciones reales		Condiciones codificadas		
W_1	W_2	X_1	X_2	Y
76	145	–1.41	0	83.3
104	145	1.41	0	81.2
90	138	0	–1.41	81.2
90	152	0	1.41	79.5
90	145	0	0	87.0
90	145	0	0	86.0

g) Represente gráficamente las corridas agregadas a la segunda región experimental, y verifique que efectivamente se completó un diseño de composición central.
h) Ajuste un modelo de segundo orden a los datos del diseño de composición central, ¿ahora sí este modelo describe la superficie? Argumente su respuesta.
i) Obtenga la gráfica de superficie de respuesta y de contornos para el modelo de segundo orden.
j) Verifique supuestos para el modelo de segundo orden.
k) Encuentre las condiciones que maximizan el rendimiento.
l) Explique en forma breve, con sus palabras, qué se hizo en las diferentes etapas para finalmente encontrar la condición óptima.

27. Para estudiar la elaboración de hojuelas de soya por extrusión, Aguilera y Kosikowski (1976) utilizaron un diseño de Box-Behnken con tres factores y tres niveles en cada factor, tal como se muestran en la siguiente tabla:

Factores	Niveles		
	–1	0	1
1. Temperatura (°C)	120	145	170
2. Contenido de humedad (%)	25	35	45
3. Velocidad de tornillo (rpm)	800	900	1 000

El diseño experimental escrito en unidades codificadas y la respuesta observada en cada punto se listan en la siguiente tabla:

x_1	x_2	x_3	Y
–1	–1	0	3.7
1	–1	0	2.0
–1	1	0	33.4
1	1	0	4.4
–1	0	–1	10.0
–1	0	1	6.0
1	0	–1	4.5
1	0	1	3.3
0	–1	–1	3.9
0	–1	1	3.7
0	1	–1	31.0
0	1	1	21.0
0	0	0	3.5
0	0	0	4.5
0	0	0	6.4

a) Ajuste un modelo de segundo orden a estos datos.
b) Verifique que el modelo sea adecuado. ¿Es útil para fines de predicción?
c) Encuentre la ecuación canónica.
d) ¿Qué tipo de superficie describe el modelo?
e) Suponga que interesa minimizar la respuesta, ¿es el tipo de superficie deseada? Si no es así, ¿qué se debe hacer?
f) De ser necesario, haga el análisis de cordillera para determinar el mejor punto posible dentro de la región experimental.

28. En un proceso de soldado de aluminio había problemas con la fuerza de unión de la soldadura. El inconveniente ocurría al soldar el poste a un tablero electrónico. Se decide

abordar el problema mediante el diseño de experimentos, utilizando como variable de respuesta la fuerza de arrastre (la fuerza necesaria para despegar la soldadura después de transcurrido un tiempo). Los factores de control que pueden influir en la fuerza de arrastre son tiempo (*T*), potencia (*P*) y fuerza (*F*). Las condiciones de operación antes del experimento eran:

Factor	Nivel	Unidades
Tiempo	6	mseg
Potencia	5	Watts
Fuerza	850	grs

Se trata de encontrar las condiciones de operación que maximizan la fuerza de arrastre. Por lo tanto, se decide utilizar en primera instancia el diseño factorial completo con dos repeticiones al centro (2^3 + *centro*) y dos réplicas. Los resultados acomodados en el orden aleatorio utilizado en cada réplica se muestran a continuación:

T	*P*	*F*	Tiempo	Potencia	Fuerza	Fza. de arrastre
0	0	0	5	4.3	750	1 981, 1 645
–1	1	–1	3	5.6	700	1 506, 1 157
1	–1	–1	7	3.0	700	1 308, 913
1	1	1	7	5.6	800	1 682, 1 049
1	–1	1	7	3.0	800	908, 1 094
–1	1	1	3	5.6	80	1 633, 1 428
1	1	–1	7	5.6	700	1 059, 1 407
–1	–1	1	3	3.0	800	875, 967
0	0	0	5	4.3	750	1 792, 1 948
–1	–1	–1	3	3.0	700	750, 666

a) ¿Qué utilidad tienen los puntos al centro?

b) Haga el análisis de estos datos. Verifique mediante la prueba de falta de ajuste (*lack of fit test*) la presencia de curvatura.

Al detectar la presencia de curvatura, el equipo de mejora decide aumentar el diseño para convertirlo en un diseño ortogonal de composición central. No se elige el rotable porque recomienda abrir los niveles de los factores más lejos de lo que permite el equipo. Los resultados observados sobre los puntos axiales, listados en orden aleatorio, se muestran a continuación:

T	*P*	*F*	Tiempo	Potencia	Fuerza	Fza. de arrastre
–1.28	0	0	2.44	4.3	750	1 404, 1 322
0	–1.28	0	4.0	2.64	750	740, 756
1.28	0	0	7.56	4.3	750	1 656, 1 512
0	0	–1.28	4.0	4.3	686	1 482, 1 485
0	1.28	0	4.0	5.96	750	1 659, 1 458
0	0	1.28	4.0	4.3	814	1 517, 1 176

c) Con estos datos complete el diseño ortogonal de composición central. Analice visualmente los datos, qué observa de destacado.

d) Ajuste un modelo de segundo orden completo.

e) Caracterice la superficie descrita por el modelo. ¿Qué tipo de superficie es?

f) Determine las condiciones de operación que maximizan la fuerza de arrastre.
g) Verifique residuos, ¿hay alguna anormalidad?

29. Se realiza el siguiente experimento para estudiar el sabor del queso panela en función de la cantidad de cuajo y sal. La variable de respuesta observada es el promedio reportado por un grupo de cinco jueces que probaron todos los quesos y los calificaron usando una escala hedónica.

Sal	Cuajo	Sabor
6	0.3	5.67
5.5	0.387	7.44
4.5	0.387	7.33
4	0.3	6.33
4.5	0.213	7.11
5.5	0.213	7.22
5	0.3	6.33
5	0.3	6.66

a) Ajuste un modelo de segundo orden a estos datos. ¿Qué porcentaje de la variación observada explica el modelo?
b) Escriba el modelo de segundo orden. Dibuje la superficie de respuesta y la gráfica de contornos.
c) Encuentre la combinación de sal y cuajo que dan el mejor sabor.

30. En el área de desarrollo de una empresa se pretende obtener un nuevo polímero de bajo peso molecular, de lograrse esto, se obtendrá un polímero que funcione como dispersante en la industria cerámica. De acuerdo con los conocimientos técnicos que se tienen, se considera que los factores críticos son: persulfato de sodio (NaPS), ácido hipofosforoso (H_3PO_2) y el isopropanol (IPA). Para encontrar las condiciones óptimas de estos tres factores, se corrió el diseño de superficie de respuesta (Box-Behnken):

Tratamiento	NaPS	H_3PO_2	IPA	Peso mol.	Viscosidad
1	0	0	0	8 392	1 075
2	–1	–1	0	9 895	2 325
3	1	–1	0	9 204	1 575
4	–1	1	0	7 882	690
5	1	1	0	7 105	420
6	–1	0	–1	8 939	1 188
7	1	0	–1	8 548	930
8	0	0	0	8 598	920
9	–1	0	1	9 152	1 275
10	1	0	1	8 992	860
11	0	–1	–1	10 504	5 600
12	0	1	–1	7 462	540
13	0	–1	1	9 368	1 225
14	0	1	1	7 772	620
15	0	0	0	8 440	1 015

a) Represente en un cubo los puntos del diseño empleado.
b) Realice un análisis estadístico completo para la variable peso molecular. Obtenga: los factores más importantes, un modelo depurado, la calidad del modelo y la verificación de supuestos.

c) ¿Cuáles son las condiciones que minimizan el peso molecular? Apóyese en gráficas de contornos y en análisis de cordillera.
d) Haga lo que se pidió en el inciso *b*), pero ahora para la variable viscosidad.
e) Encuentre una condición que de manera simultánea minimice ambas respuestas. Hágalo gráficamente.

31. Los datos que se presentan en la siguiente tabla fueron recopilados en un experimento para optimizar el crecimiento de cristales en función de tres variables x_1, x_2, x_3. Se buscan valores altos de Y (rendimiento en gramos).

x_1	x_2	x_3	Y
–1	–1	–1	66
1	–1	–1	70
–1	1	–1	78
1	1	–1	60
–1	–1	1	80
1	–1	1	70
–1	1	1	100
1	1	1	75
0	0	–1.68	100
0	0	1.68	80
0	–1.68	0	68
0	1.68	0	63
0	0	–1.68	65
0	0	1.68	82
0	0	0	113
0	0	0	100
0	0	0	118
0	0	0	88
0	0	0	100
0	0	0	85

a) ¿Qué diseño se empleó?
b) Ajuste un modelo de segundo orden y determine los componentes significativos y más importantes.
c) ¿El modelo describe adecuadamente la superficie? Argumente.
d) Considere las dos variables independientes más importantes en el modelo y obtenga la superficie de respuesta y de contornos en función de éstas.
e) ¿Con la superficie que encontró en el inciso anterior se puede hacer algún tipo de inferencia? Argumente.
f) Verifique los supuestos para el modelo. ¿Observa algo relevante?
g) Observe las respuestas obtenidas en los puntos centrales del diseño. ¿Esto le ayuda a explicar lo que ocurrió? Argumente.
h) ¿A qué cree que se debe el mal ajuste del modelo?

32. Supongamos el modelo ajustado dado por:

$$\hat{Y} = 10.46 - 0.57x_1 + 0.18x_2 + 0.45x_3 - 0.67x_1^2 + 0.56x_2^2 - 0.27x_3^2 - 0.67x_1x_2 + 1.18x_1x_3 + 0.23x_2x_3$$

a) Encuentre el punto estacionario y diga de qué tipo es.
b) Si se obtiene un punto silla, realice un análisis de cordillera.

33. Considere el modelo ajustado de segundo orden dado por:

$$\hat{Y} = 8.25 + 0.0396x_1 + 0.1884x_2 - 0.4313x_1^2 - 0.3563x_2^2 + 0.3x_1x_2$$

a) Dibuje la superficie descrita por el modelo.
b) ¿Qué tipo de superficie es?
c) Localice las coordenadas del punto estacionario.
d) Encuentre un punto sobre la dirección de mínima pérdida a partir del punto estacionario.

34. Determine la trayectoria de ascenso máximo dada por el modelo $\hat{Y} = 60 + 1.5x_1 + 2x_3$. Encuentre cinco puntos sobre ella, utilice un paso unitario en la variable de mayor influencia.

Capítulo 13
Optimización simultánea de varias respuestas

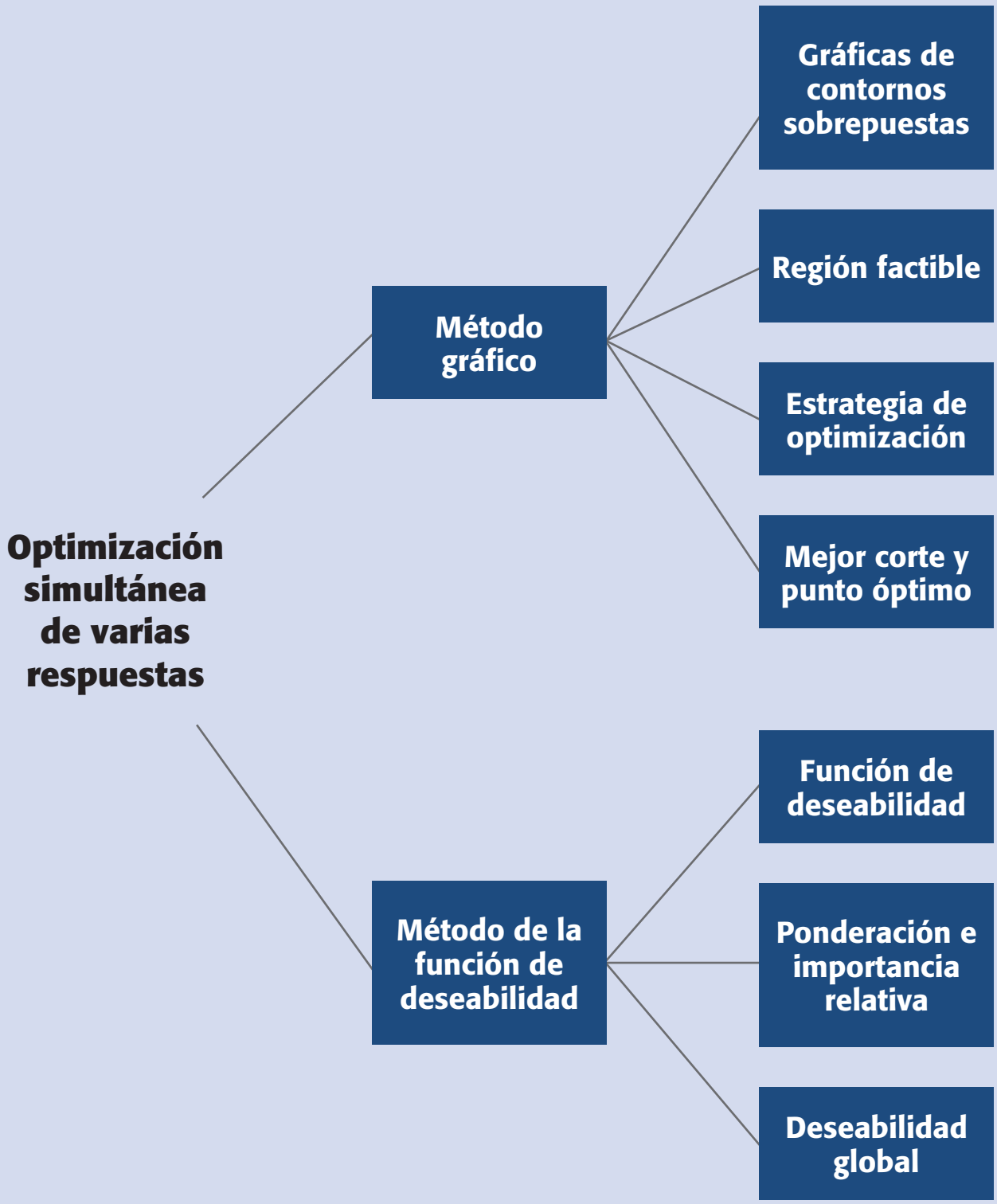
Sumario

- Optimización simultánea
- Método gráfico
- Método de la función de deseabilidad
- Trabajar con un software estadístico

Objetivos de aprendizaje

- Comprender la problemática que implica la optimización simultánea de varias respuestas, así como su importancia considerando que todos los productos y procesos tienen varias características de interés que determinan su calidad.
- Aplicar el método gráfico para optimizar varias respuestas.
- Aplicar el método de la función de deseabilidad en problemas de respuesta múltiple.

Mapa conceptual



Conceptos clave

- Deseabilidad global
- Exponentes *s* y *t*
- Función de deseabilidad
- Gráfica de contornos
- Hipersuperficie
- Mejor corte
- Método gráfico
- Modelo jerárquico
- Óptimo individual
- Óptimo simultáneo
- Pesos w_i (importancia relativa)
- Punto óptimo gráfico
- Región factible
- Superposición de superficies

Óptimo individual
Combinación de factores o tratamiento donde se predice el mejor valor posible de una respuesta dentro de la región experimental.

Óptimo simultáneo
Combinación de los factores de control donde todas las respuestas de interés toman valores aceptables. Es una solución de compromiso.

Optimización simultánea

En el capítulo anterior hicimos énfasis en encontrar el punto óptimo considerando sólo una variable de respuesta (característica de calidad). Sin embargo, es típico considerar diversas características (variables) para lograr productos con mejor calidad y propiedades. Por ejemplo, un alimento tiene varias propiedades como: textura, pH, color, sabor, apariencia, etc., y todas son importantes para que el alimento sea bien aceptado por los consumidores. Si la optimización sólo se hace para una característica del producto podrían resultar condiciones inadecuadas para las otras características. Por ello es imprescindible contar con técnicas que sirvan para que, en la medida de lo posible, se optimicen simultáneamente todas las respuestas de interés.

El problema de la *optimización simultánea* radica en que, por lo general, los óptimos individuales no son las mismas combinaciones de los factores de control ($X_1, X_2,\ldots, X_k$). Esto hace necesario buscar una *solución compromiso*, donde todas las variables tengan un nivel satisfactorio. A esa solución compromiso la llamaremos *óptimo simultáneo*. Por ejemplo, en la figura 13.1 se muestran las superficies de dos respuestas en una misma región experimental. Si en ambas respuestas interesara el mínimo, es claro que los óptimos individuales se encuentran en extremos opuestos de la región experimental. Por lo tanto, habrá que buscar otra combinación (punto) donde las dos variables sean al mismo tiempo lo menor posible. En la figura 13.1 se muestra dónde está tal óptimo simultáneo.

En este capítulo presentaremos dos métodos de optimización simultánea, uno gráfico (De la Vara y Domínguez, 2002) y el otro analítico, basado en una función de deseabilidad (Derringer y Suich, 1980). Ambos métodos proveen soluciones consistentes tanto con los datos observados en el experimento como con la información que se les provee. Asimismo, son intuitivos y flexibles en el sentido de que permiten

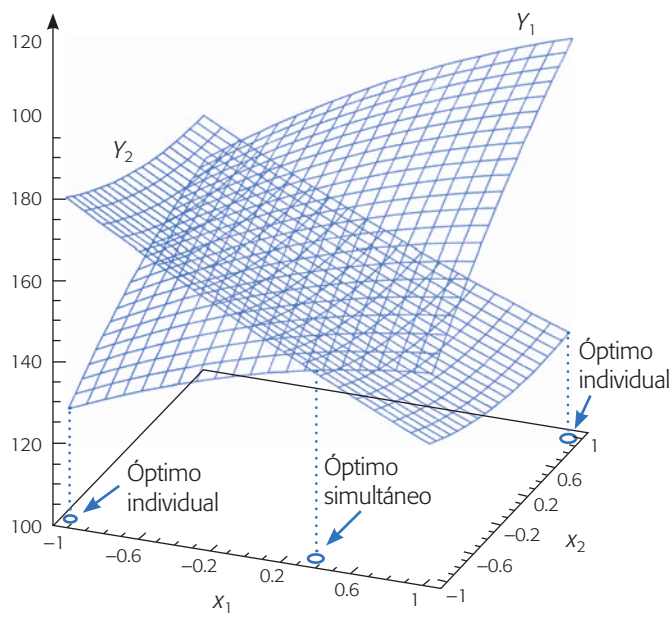


Figura 13.1 El óptimo simultáneo (global) es una solución de compromiso.

balancear de diversas formas la importancia relativa de las respuestas. Para aplicar estos métodos es importante contar con el software apropiado (véase última sección de este capítulo).

Para ambos métodos se supone que cada una de las m respuestas (Y_1, Y_2,..., Y_m) a optimizar, está modelada adecuadamente por un modelo de segundo orden en términos de los mismos k factores de control (X_1, X_2,…, X_k). Es decir, para empezar la optimización simultánea se deben tener los m modelos estimados, dados por:

$$\hat{Y}_l = \hat{\beta}_{0l} + \sum_{i=1}^{k} \hat{\beta}_{il} x_i + \sum_{i=1}^{k} \hat{\beta}_{iil} x_i^2 + \sum_{i=1<j=1}^{k} \sum_{j=1}^{k} \hat{\beta}_{ijl} x_i x_j;\ l = 1, 2, \ldots, m \qquad \textbf{(13.1)}$$

No necesariamente tienen que ser modelos cuadráticos completos, pero se recomienda que sean *modelos jerárquicos*, es decir, que por cada interacción o término cuadrático en el modelo, éste también incluya los términos más simples que se pueden formar con los factores involucrados. Es preciso verificar que cada modelo cumpla los supuestos tradicionales de normalidad, la varianza constante y la independencia de los residuos, y que el coeficiente de determinación R^2_{aj} de cada modelo sea de al menos 70%. Tanto el método gráfico como el método de la función de deseabilidad requieren del conocimiento de especificaciones para cada una de las variables de respuesta.

Modelo jerárquico
Si un efecto de interacción está presente en el modelo, entonces se debe incluir en el modelo los términos más simples que forman dicha interacción.

Ejemplo 13.1 (Optimización de neumáticos)

En Derringer y Suich (1980) se presenta un problema que consiste en encontrar la combinación óptima de tres ingredientes de un compuesto de las bandas para neumáticos: silicato (X_1), silanio (X_2) y sulfuro (X_3), considerando de manera simultánea cuatro variables de respuesta de interés, cuyos nombres y especificaciones son:

$$Y_1 > 120 = EI;\ \textit{índice de abrasión}$$
$$Y_2 > 1\,000 = EI;\ \textit{módulo}\ 200\%$$
$$EI = 400 < Y_3 < 600 = ES;\ \textit{elongación}$$
$$EI = 60 < Y_4 < 75 = ES;\ \textit{dureza}$$

Si bien, las dos primeras variables no tienen límite de especificación superior, y son del tipo *entre más grande mejor*, desde el punto de vista práctico se considera que no hay ninguna ganancia adicional en estas propiedades cuando Y_1 y Y_2 toman valores mayores a 170 y 1 300, respectivamente. En este sentido, estos últimos números se consideran los valores objetivos de dichas variables. Por su parte, las variables Y_3 y Y_4 toman el punto central de las especificaciones como valores objetivo. Para buscar la optimización de estas cuatro variables se corrió el diseño de composición central de la tabla 13.1. Al ajustar a cada variable de respuesta un modelo de segundo orden se obtiene que:

$$\hat{Y}_1 = 139.1 + 16.5x_1 + 17.9x_2 + 10.9x_3 - 4.0x_1^2 - 3.45x_2^2 - 1.57x_3^2 + 5.13x_1x_2 + 7.13x_1x_3 + 7.88x_2x_3$$
$$\hat{Y}_2 = 1261.1 + 268.2x_1 + 246.5x_2 - 122.5x_3 - 83.57x_1^2 - 124.82x_2^2 + 199.17x_3^2 + 94.13x_1x_3 + 436.6x_1^2x_3$$
$$\hat{Y}_3 = 400.4 - 99.7x_1 - 31.4x_2 - 73.9x_3 + 7.93x_1^2 + 17.31x_2^2 + 0.43x_3^2 + 8.75x_1x_2 + 6.25x_1x_3 + 1.25x_2x_3$$
$$\hat{Y}_4 = 68.9 - 1.4x_1 + 4.3x_2 + 1.6x_3 + 1.56x_1^2 + 0.06x_2^2 - 0.32x_3^2 - 1.63x_1x_2 + 0.13x_1x_3 - 0.25x_2x_3$$

Note que el modelo para Y_2 tiene el término cúbico $x_1^2x_3$, con lo cual se logra mejorarlo un poco con respecto al modelo de segundo orden utilizado por Derringer y Suich (1980).

El coeficiente de determinación y algunos otros elementos para evaluar la calidad de ajuste de cada modelo se muestran en la tabla 13.2.

Todos los modelos tienen un ajuste satisfactorio; sin embargo, el modelo 2 tiene un ajuste apenas satisfactorio. En cuanto al rango de predicción, éste se refiere a los valores de cada Y que podrían predecir cada modelo. Ahí se observa que los rangos deseados para cada Y están inmersos en el rango de predicción, que en principio hace factible la obtención de un óptimo simultáneo. Para ello aplicaremos el método gráfico y el de la función de deseabilidad. En caso de que en este rango no se incluyan los valores factibles, es probable que experimentalmente tampoco se hayan observado valores factibles para tal variable; por lo tanto, habría que resolver antes este problema, o bien, reconsiderar las especificaciones.

Método gráfico

Quizás lo primero que se ocurre al optimizar varias respuestas es superponer sobre la región experimental las m superficies de respuesta descritas por los m modelos ajustados, y localizar dentro de ella subregiones en las cuales todos los modelos predicen valores aceptables para las respuestas. Cuando sólo se tienen dos factores de

Tabla 13.1 Diseño y resultados experimentales para el ejemplo 13.1.

Corrida	X_1	X_2	X_3	Y_1	Y_2	Y_3	Y_4
1	–1	–1	1	102	900	470	67.5
2	1	–1	–1	120	860	410	65
3	–1	1	–1	117	800	570	77.5
4	1	1	1	198	2 294	240	74.5
5	–1	–1	–1	103	490	640	62.5
6	1	–1	1	132	1 289	270	67
7	–1	1	1	132	1 270	410	78
8	1	1	–1	139	1 090	380	70
9	–1.63	0	0	102	770	590	76
10	1.63	0	0	154	1 690	260	70
11	0	–1.63	0	96	700	520	63
12	0	1.63	0	163	1 540	380	75
13	0	0	–1.63	116	2 184	520	65
14	0	0	1.63	153	1 784	290	71
15	0	0	0	133	1 300	380	70
16	0	0	0	133	1 300	380	68.5
17	0	0	0	140	1 145	430	68
18	0	0	0	142	1 090	430	68
19	0	0	0	145	1 260	390	69
20	0	0	0	142	1 344	390	70

Tabla 13.2 Calidad de ajuste de cada modelo, ejemplo 13.1.

Modelo	R^2_{aj}	Error estándar: $\sqrt{CM_E}$	Rango de predicción
*M*1	0.95	5.61	96-198
*M*2	0.75	232.78	490-2 294
*M*3	.96	20.55	240-640
*M*4	0.92	1.27	62.5-78

control (X_1 y X_2) es bastante fácil superponer las superficies de respuesta en su forma de *contornos* sobre la región experimental, que en este caso se puede dibujar como un cuadrado o un círculo centrado en el origen. Pero cuando se tienen más de dos factores, como es el caso de los neumáticos, las superficies no se pueden dibujar de una sola vez sobre toda la región experimental. Por ello, se tendrán que graficar tomando dos factores a la vez y fijando a los restantes.

Cuando los factores de control son tres, la región experimental es un cubo (o esfera) centrado en el origen y se puede dibujar una *superposición* de las k superficies de contornos sobre cada corte del cubo, lo cual implica fijar cada vez uno de los factores de control. En la figura 13.2 se grafica en el sentido del plano (X_1, X_3) y el factor X_2 se fija en tres posibles cortes ($x_2 = -1$; $x_2 = 0$; $x_2 = 1$). Los contornos que vemos sobre cada lámina o corte son sólo un aspecto o visión de los contornos originales tridimensionales de la *hipersuperficie* de respuesta para una Y. Con tres factores, el cubo se podría rebanar en el sentido de los planos (X_1, X_2), (X_1, X_3) y (X_2, X_3). En cada una de estas formas de rebanar al cubo, la ubicación específica de un corte está dada por el valor en que se fija el tercer factor. Se recomienda que los cortes se hagan en el sentido de los factores con menor influencia en todos los modelos, ya que si se usan unidades codificadas, en la mayoría de los modelos ajustados son identificadas en valor absoluto por los coeficientes más pequeños. De esta manera se apreciará mejor el efecto del factor de mayor influencia.

En la figura 13.2 sólo se muestran cortes (planos) de una superficie de respuesta. Para las otras Y sus correspondientes planos se superponen y, de esa manera, se va identificando las regiones factibles (el auxilio de un software es indispensable).

A continuación explicaremos de manera detallada el *método gráfico* para el caso de $k = 3$ factores y lo ilustramos con el ejemplo 13.1. Como veremos, este método gráfico de optimización consiste en fijar restricciones para cada Y, y determinar el mejor corte que hace que se cumpla con tales restricciones. Enseguida se estrechan las restricciones y se vuelve a determinar el mejor corte, hasta que la región factible sea pequeña.

Paso 1. Para cada variable de respuesta y considerando los cortes del cubo en el sentido de los dos factores con menor influencia en todos los modelos, en cada corte se dibujan las dos curvas de nivel que corresponden a las especificaciones. En el caso del ejemplo 13.1 se decide hacer los cortes en el sentido del plano (X_1, X_2), a la altura de $x_3 = -1$, 0, 1. En cada corte se dibujan las curvas de nivel dadas por Y_1(120, 170), Y_2 (1 000, 1 300), Y_3 (400, 600) y Y_4 (60, 75). El resultado se muestra en la figura 13.3. El software señala (dejando sin sombrear o pintando de otro color más claro) la subregiones donde todas las respuestas predicen valores factibles. En la fi-

Gráfica de contornos
Superficie de respuesta representada mediante curvas de nivel.

Superposición de superficies
Consiste en colocar las diferentes superficies individuales sobre la misma región experimental en función de dos factores.

Hipersuperficie
Superficie de respuesta que se encuentra en más de tres dimensiones porque se tienen tres o más factores. Sólo se puede ver parcialmente en cortes bidimensionales.

Método gráfico
Consiste en superponer dos curvas de nivel para cada variable de respuesta e identificar gráficamente regiones factibles donde todas las respuestas cumplen los requerimientos.

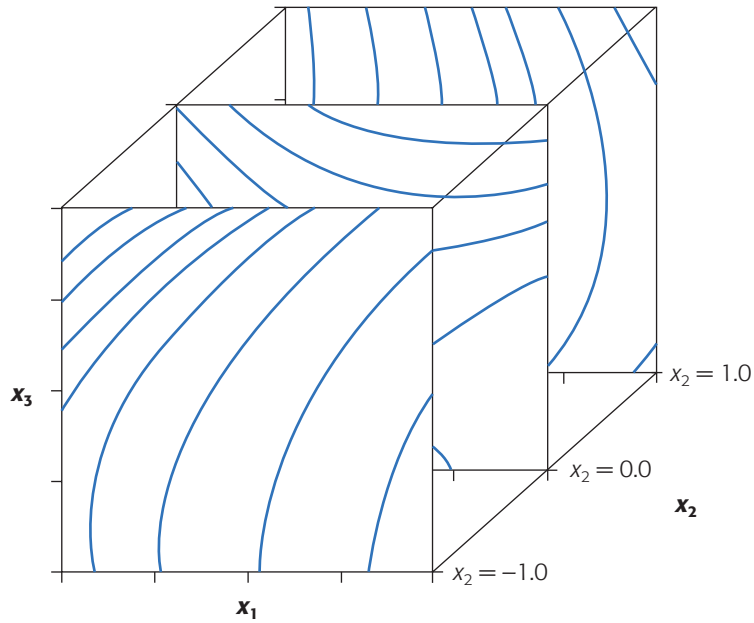


Figura 13.2 Contornos de una hipersuperficie en sentido (X_1, X_3) con cortes en $x_2 = -1, 0, 1$.

gura se observa que existe región factible en todos los cortes a lo largo del eje X_3. Nótese que aunque Y_1 y Y_2 son variables de respuesta con especificación unilateral, también se han dibujado dos curvas de nivel: una que corresponde a la especificación y otra seleccionada como un límite práctico (valor objetivo) a partir del cual no habría ganancia adicional en la calidad de la respuesta.

Región factible
Zona de la región experimental donde todas las respuestas cumplen con los requerimientos establecidos.

Se elige como el mejor corte aquel que maximiza el tamaño de la *región factible*. Existen varias posibilidades en este primer paso:

a) Cuando en casi cualquier corte existe una región factible grande, es indicio de que las restricciones impuestas se pueden cumplir fácilmente y lo que procede es estrechar las especificaciones para reducir el tamaño de la subregión factible.
b) Pero, si por el contrario, no se encuentra ningún corte donde exista región factible, entonces el problema no tiene solución para esas restricciones y significa que al menos una de las variables tendría que tomar valores fuera de especificaciones para que las restantes sí puedan cumplir con las suyas.
c) Una tercera posibilidad es que se descubran varias regiones factibles separadas, en cuyo caso es probable que la más grande de ellas arroje los mejores resultados; sin embargo, también es posible hacer consideraciones de costos de operación y de variabilidad para elegir una de las regiones.

En el ejemplo, dado lo anterior y la figura 13.3, se está en el caso de la opción *a*), por lo que iremos directo al siguiente paso.

Paso 2. Si ya se ubicó una región factible o se está en el caso de la opción *a*) anterior, lo que procede es ir estrechando el rango de las respuestas hacia sus valores más

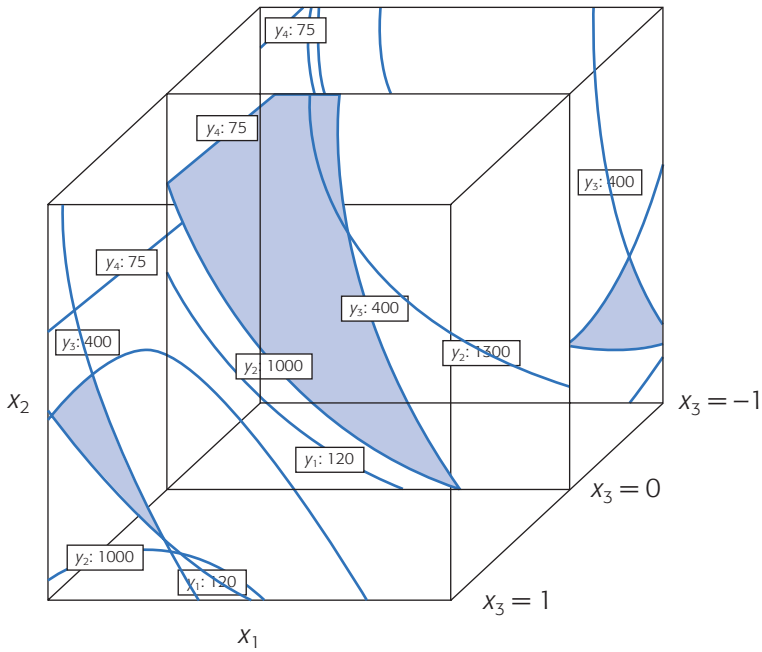


Figura 13.3 Primer paso del método gráfico: en todos los cortes hay región factible.

deseables. Para cada selección de nuevas curvas de nivel se vuelve a determinar el *mejor corte* posible, y así se sigue hasta lograr una región factible suficientemente pequeña que, dependiendo de la fineza deseada, no mida más de 0.1 unidades codificadas. Note que cada vez que se estrecha el rango las curvas de nivel son más cercanas a los valores deseables o valores nominales de las respuestas. Habría al menos dos estrategias o criterios para ir realizando el estrechamiento del rango hacia los valores más deseables de las respuestas.

Mejor corte
Es aquel donde la región factible maximiza su tamaño. Puntos sobre este corte predicen valores de las respuestas que cumplen de mejor manera los límites dados.

a) *Criterio proporcional al rango*. Considerando variables igualmente importantes, el acercamiento a los valores más deseables se hace en la misma proporción, o bien, si existen variables más importantes, éstas se estrechan en mayor proporción. Este criterio no funciona bien si los errores estándar ($\sqrt{CM_E}$) de los modelos ajustados son muy diferentes entre sí, lo cual ocurre con frecuencia si las variables se miden en escalas muy diferentes.

b) *Criterio proporcional al error estándar*. En cada ciclo los estrechamientos del rango en cada respuesta se hacen de manera proporcional al tamaño del error estándar estimado para cada modelo. Idealmente se buscaría alejar los valores predichos de las respuestas al menos dos errores estándar de los límites de especificación.

En el ejemplo, dado que en cualquier corte hay una región factible, se procede a estrechar las especificaciones para cada Y de manera proporcional al error estándar de su modelo. Se hacen estas reducciones y se logra llegar hasta una reducción de alrededor de 2.5 errores estándar en las respuestas Y_1, Y_3 y Y_4. En la respuesta Y_2 no se logra este nivel de reducción, ya que tiene un error estándar grande con respecto al límite práctico de 1 300 definido para esta respuesta. Así, con esta variable sólo se

logra ir hasta 1.27 veces el error estándar. Estos resultados se muestran en la figura 13.4, donde se observa una pequeña región factible.

Paso 3. Una vez que se tiene una región factible suficientemente pequeña, ya no se mueven las especificaciones de las Y y se procede a ubicar el mejor corte, variando el factor que no está representado en el plano de corte. El *mejor corte* es aquel que hace que la subregión se observe lo más grande posible para las condiciones dadas. Por último, se proyecta el centro de la región en su mejor corte y se determinan las coordenadas del punto óptimo simultáneo.

En el ejemplo, de acuerdo con la figura 13.4, el mejor corte es en $x_3 = -0.38$ y de la proyección sobre los ejes x_1 y x_2 se obtiene el siguiente punto óptimo simultáneo:

Óptimo simultáneo			Predichos			
x_1	x_2	x_3	Y_1	Y_2	Y_3	Y_4
–0.36	0.42	–0.38	134.26	1 303.4	454.13	71.10

Tres de los valores predichos se alejan de la especificación más cercana al menos 2.5 veces, mientras que el error estándar y el segundo de ellos sólo se aleja 1.27 veces del error estándar. Es decir, dado su peor modelo, la respuesta Y_2 es una de las más afectadas. Si se quisiera mejorar más alguna de las respuestas necesariamente sería en detrimento de al menos una de las respuestas restantes. Cabe mencionar que fue necesario relajar (aumentar) a 1 320 la curva del nivel superior de Y_2 para obtener mejores valores estimados de la misma, lo cual no es problema porque Y_2 es del tipo "mientras más grande mejor". Es importante señalar que la última región factible no debe quedar delimitada por algún límite práctico puesto que éstos son un tanto arbitrarios y siempre se podrían relajar más hacia mejores, como se hizo con el límite superior de la respuesta Y_2 que quedó en 1 320 y no delimita la región óptima (figura 13.4).

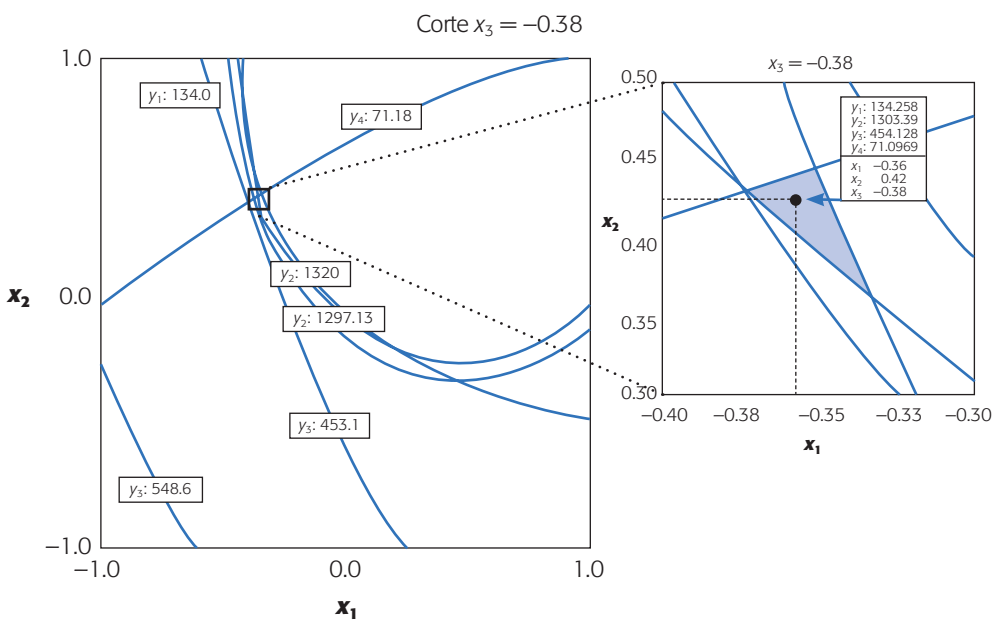


Figura 13.4 Punto óptimo simultáneo, método gráfico.

Algunas ventajas del método gráfico con respecto a los métodos numéricos son: que permite explorar directamente la superficie de respuesta simultánea, y de esta manera el experimentador obtiene una visión gráfica del comportamiento relativo de las respuestas. Además, el método asegura que el punto óptimo simultáneo encontrado sea en realidad un óptimo global, ya que el método puede detectar la presencia de varias regiones factibles separadas que pueden dar lugar a óptimos locales. Obviamente, no sólo localiza un *punto óptimo*, más aún localiza zonas o regiones óptimas. Entre sus desventajas se tiene que a simple vista puede parecer difícil; sin embargo, no lo es cuando se cuenta con un buen graficador. También se vuelve impráctico cuando se tienen cinco o más factores, debido a la cantidad de gráficas que será necesario ver.

Punto óptimo gráfico
Centro de la pequeña región factible a la que se llega después de ir estrechando hacia los valores ideales las especificaciones originales de las respuestas.

Método de la función de deseabilidad

Este método fue propuesto originalmente por Harrington (1965) y después fue mejorado por Derringer y Suich (1980) y Derringer (1994). Consiste en definir una función en el espacio de factores que estima la deseabilidad global (DG) del producto en cada punto; de esta forma, convierte el problema de optimización multivariado en un problema de optimización univariado. Basta maximizar DG para obtener el punto óptimo buscado. Para definir la DG se requiere que todas las Y estén en la misma escala, y esto se logra transformando cada respuesta predicha $\hat{Y}_i(\mathbf{x})$ en un valor de deseabilidad individual $d_i(\mathbf{x})$ que cae en el intervalo [0, 1]. De esta manera, $d_i(\mathbf{x})$ mide la deseabilidad del punto $\mathbf{x} = (x_1, x_2,..., x_k)$ con respecto a la variable Y_i. La transformación $d_i(\mathbf{x})$ se hace en términos de las especificaciones y del valor objetivo de cada Y.

Función de deseabilidad
Función de los factores de control que transforma los valores predichos por el modelo a una escala [0, 1], para indicar qué tan deseables son.

En particular, si la variable Y_i tiene por especificaciones inferior y superior a EI_i y ES_i, y su valor objetivo o nominal es T_i, se define la transformación d_i como:

$$d_i(\mathbf{x}) = \begin{cases} \left[\dfrac{\hat{Y}_i(\mathbf{x}) - EI_i}{T_i - EI_i}\right]^s & \text{si} \quad EI_i \le \hat{Y}_i(\mathbf{x}) \le T_i \\ \left[\dfrac{\hat{Y}_i(\mathbf{x}) - ES_i}{T_i - ES_i}\right]^t & \text{si} \quad T_i < \hat{Y}_i(\mathbf{x}) \le ES_i \\ 0 \quad \text{si} \quad \hat{Y}_i(\mathbf{x}) < EI_i \ \text{ o } \ \hat{Y}_i(\mathbf{x}) > ES_i \end{cases} \tag{13.2}$$

donde s y t son exponentes que sirven para elegir la forma deseada de la transformación y con ello reflejan los deseos del experimentador sobre cada Y. Se toman valores grandes (digamos $s, t \ge 10$) cuando se quiere que la deseabilidad d_i sólo tome valores grandes cuando $\hat{Y}_i$ cae cerca de su valor objetivo. Se toman valores pequeños para s y t ($s, t \le 0.1$) si se quiere que cualquier valor de $\hat{Y}_i$ dentro del intervalo $[EI_i, ES_i]$ sea igualmente deseable (véase figura 13.5). El valor por omisión de estos exponen-

tes es 1, lo cual sugiere un incremento lineal de la deseabilidad hacia el valor objetivo. Si la respuesta tiene especificaciones de un solo lado, lo que se hace es tomar el valor objetivo (T_i) igual al valor a partir del cual se considera que no hay ganancia adicional en la calidad de la respuesta. Es decir, en la función dada por la ecuación (13.2) desaparece una de las restricciones y la figura 13.5 se reduce a uno de los lados con respecto al valor objetivo T_i, véase ecuación (13.5) más adelante.

Deseabilidad global
Mide la deseabilidad promedio de todas las respuestas en cada combinación de los factores. Se obtiene con la media geométrica de las deseabilidades individuales.

Una vez calculadas las m deseabilidades individuales sobre el punto $\mathbf{x}$, la *deseabilidad global* (DG) de $\mathbf{x}$ es definida por la media geométrica ponderada:

$$\mathrm{DG}(\mathbf{x}) = \left(d_1^{w_1} \times d_2^{w_2} \times \cdots \times d_m^{w_m}\right)^{1/\sum w_i} \qquad \textbf{(13.3)}$$

donde los pesos w_i son constantes que permiten balancear la importancia relativa de cada variable; mientras más grande es el peso dado a una variable con respecto a las restantes, más grande será la exigencia para que el punto óptimo global beneficie a tal variable. Si todas son igualmente importantes, $w_i = 1$ para $i = 1, 2, \ldots, m$, y la DG toma la forma siguiente:

$$\mathrm{DG} = \sqrt[n]{d_1 \times d_2 \times \cdots \times d_n} = \left(d_1 \times d_2 \times \cdots \times d_n\right)^{1/n} \qquad \textbf{(13.4)}$$

Si d_i es igual a 1 significa que la correspondiente respuesta predicha $\hat{Y}_i$ toma su valor máximo deseable. Si $d_i = 0$ la respuesta $\hat{Y}_i$ predice un valor inaceptable y en este caso la deseabilidad global es cero (DG = 0), lo cual significa que todo el producto es inaceptable, independientemente de los valores de las respuestas restantes. Esto último explica el uso de la media geométrica en la definición de la deseabilidad global DG.

Exponentes *s* y *t*
Ponderaciones que permiten definir la forma de la función de deseabilidad para cada respuesta.

Note que los exponentes s y t definidos en la transformación (13.2) se pueden introducir como parte de los pesos w_i, pero es importante elegir los exponentes y los pesos de manera separada, ya que los primeros definen la forma de la función de deseabilidad que se quiere para cada respuesta individual y los segundos precisan la importancia relativa entre las respuestas. El punto óptimo simultáneo es el punto $\mathbf{x}_0 = (x_{10}, x_{20}, \ldots, x_{k0})$ sobre el cual la función DG($\mathbf{x}_0$) es máxima. Para encontrar este máximo se recurre a algún método numérico.

Pesos w_i (importancia relativa)
Ponderaciones que surgen de considerar qué tan importante es cada respuesta comparada con las demás.

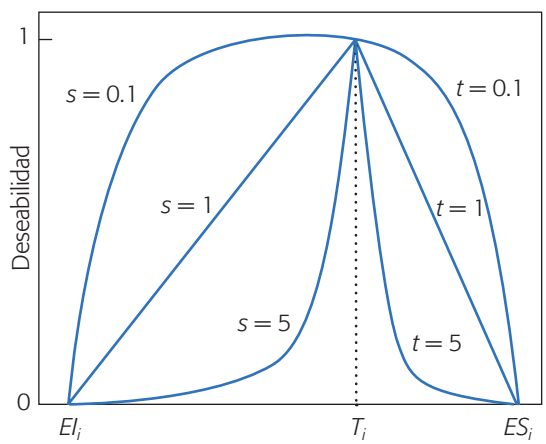


Figura 13.5 Función de los exponentes para la obtención de d_i (deseabilidad).

Aplicar lo anterior al problema del ejemplo 13.1 (de los neumáticos) es directo, ya que se tienen los cuatro modelos ajustados y las especificaciones para las cuatro variables de respuesta. Por ello, empleando la expresión (13.2) se obtienen las cuatro funciones de deseabilidad individuales con exponentes unitarios y pesos iguales. Por ejemplo, para la Y_1 la transformación d_1 está dada por:

$$d_1(\mathbf{x}) = \begin{cases} \left[\dfrac{\hat{Y}_1(\mathbf{x}) - 120}{170 - 120}\right] & \text{si} \quad 120 \le \hat{Y}_1(\mathbf{x}) \le 170 \\ 1 \quad \text{si} \quad \hat{Y}_1(\mathbf{x}) > 170 \\ 0 \quad \text{si} \quad \hat{Y}_1(\mathbf{x}) < 120 \end{cases} \tag{13.5}$$

Si se hace algo similar para las otras variables, la DG se consigue con la ecuación (13.4). Al maximizar esta función se obtiene que la solución óptima es el punto $\mathbf{x}_0 =$ (–0.10, 0.09, –0.79), sobre el cual se predicen los valores $\hat{Y}(\mathbf{x}_0)$ = (129.37, 1 478.53, 466.78, 68.01), cuyas deseabilidades individuales son $d[\hat{Y}(\mathbf{x}_0)]$ = (0.187, 1.000, 0.668, 0.932), respectivamente. Nótese que la variable más sacrificada es Y_1, al encontrarse a un nivel de 18.7% de deseabilidad frente a la Y_2 que se predice en 100% de su deseabilidad. Esto se puede modificar aumentando los pesos y/o los exponentes para Y_1.

En la figura 13.6 se dibuja la superficie de respuesta para la DG sobre uno de los planos que pasan por el punto óptimo. La deseabilidad global sobre el punto óptimo resulta ser de 0.584, que es aceptable y, por definición, resulta de la media geométrica de las deseabilidades individuales.

Una desventaja tanto del método gráfico como del método de la función de deseabilidad es que no consideran la aleatoriedad de los predichos $\hat{Y}_i$, lo cual puede causar que dos respuestas con deseabilidades similares no predigan la misma cantidad de producto fuera de especificaciones en los intervalos de predicción correspondientes. Por ejemplo, 0.187 de deseabilidad de la variable $\hat{Y}_1$ es sólo el valor promedio,

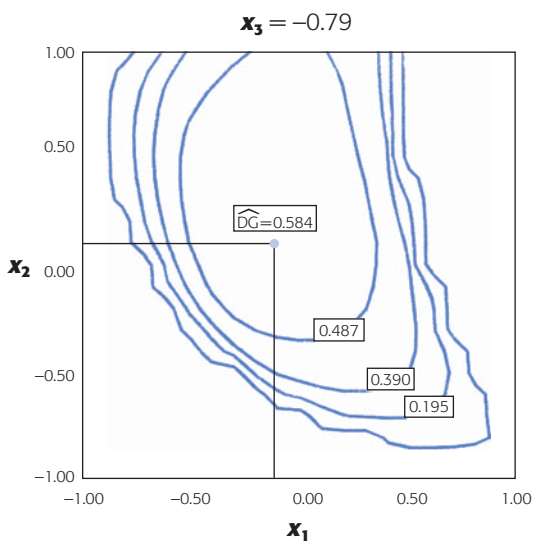


Figura 13.6 Función de deseabilidad global y punto óptimo simultáneo.

y si se toma en cuenta que el intervalo de predicción para la observación futura sobre el punto óptimo (véase capítulo 11) está dado por [115.84, 142.90]. Entonces, en la práctica se podrían tener algunos valores de Y_1 inferiores a 120, lo cual no se desea.

Una manera de mejorar un poco esta situación es mediante otras ponderaciones diferentes a la unitaria. Por ejemplo, al ponderar con peso 1.3 la primera variable, la solución resultante predice los valores $\hat{Y}_i\,(x_0)$ = (131.4, 1 481.36, 457.97, 68.56), lo que implica una mejor deseabilidad para Y_1 a costa de empeorar un poco los valores de Y_3 y Y_4. En resumen, las dos soluciones encontradas con el método de la función de deseabilidad usando dos ponderaciones (*weights*) diferentes, son:

Solución	x_{10}	x_{20}	x_{30}	$\hat{Y}_1\,(x_0)$	$\hat{Y}_2\,(x_0)$	$\hat{Y}_3\,(x_0)$	$\hat{Y}_4\,(x_0)$
DS1	–0.10	0.09	–0.79	129.37	1 478.53	466.78	68.01
DS2	–0.08	0.19	–0.73	131.4	1 481.36	457.97	68.56

Las soluciones del método de la función de deseabilidad son mejores que las obtenidas por el método gráfico, en principio porque se relajó el valor objetivo de la segunda variable hasta 1 478.53, que es del tipo "mientras más grande mejor". Es posible mostrar (véase ejercicio 12) que con el método gráfico, usando un valor objetivo de 1 500 para la variable Y_2, y aplicando el criterio de reducción proporcional al error estándar del modelo, se llega a una solución similar a la DS2.

Ejemplo 13.2 Optimización de un proceso de soldadura en semiconductores

En la manufactura de semiconductores se ensambla un módulo híbrido en un paquete de premoldeado, en el cual se unen alambres entre las guías (posición A) y los chips de silicón (posición B) (Del Castillo, E. *et al*., 1996). La unión se logra por medio de temperatura; los paquetes pasan por un bloque caliente después de recibir un flujo de nitrógeno también caliente. Las variables de respuesta de interés son: las temperaturas inicial, final y máxima alcanzadas en las posiciones A y B, con especificaciones y valores objetivos dados por:

Variables de respuesta	EI_i	T_i	ES_i
Y_1 = temperatura máxima en la posición A.	185	190	195
Y_2 = temperatura en la posición A cuando empieza el enlace.	170	185	195
Y_3 = temperatura en la posición A cuando se termina el enlace.	170	185	195
Y_4 = temperatura máxima en la posición B.	185	190	195
Y_5 = temperatura en la posición B cuando empieza el enlace.	170	185	195
Y_6 = temperatura en la posición B cuando termina el enlace.	170	185	195

Se supone que las Y dependen de tres factores: tasa de flujo (x_1), temperatura de flujo (x_2) y temperatura de bloque (x_3). Los resultados observados en un diseño Box-Behnken se muestran en la tabla 13.3. Los mejores modelos ajustados que describen el comportamiento de las respuestas se muestran en la tabla 13.4. Para cada modelo se incluye su error estándar. Por ejemplo el modelo para Y_1 está dado por

$$\hat{Y}_1 = 174.9 + 23.37x_2 + 3.63x_3 - 19.0x_2x_3$$

Tabla 13.3 Experimento sobre soldadura en semiconductores.

No. de exp.	X_1	X_2	X_3	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
1	–1	–1	0	139	103	110	110	113	126
2	1	–1	0	140	125	126	117	114	131
3	–1	1	0	184	151	133	147	140	147
4	1	1	0	210	176	169	199	169	171
5	–1	0	–1	182	130	122	134	118	115
6	1	0	–1	170	130	122	134	118	115
7	–1	0	1	175	151	153	143	146	164
8	1	0	1	180	152	154	152	150	171
9	0	–1	–1	132	108	103	111	101	101
10	0	1	–1	206	143	138	176	141	135
11	0	–1	1	183	141	157	131	139	160
12	0	1	1	181	180	184	192	175	190
13	0	0	0	172	135	133	155	138	145
14	0	0	0	190	149	145	161	141	149
15	0	0	0	180	141	139	158	140	148

Tabla 13.4 Modelos ajustados para los datos de la tabla 13.3.

Coeficiente	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
$\hat{\beta}_0$ (const.)	174.93	141.00	139.20	154.86	139.29	146.86
$\hat{\beta}_1$ (x_1)		6.00	6.63	8.50	4.25	4.50
$\hat{\beta}_2$ (x_2)	23.37	21.62	16.00	30.63	19.75	15.62
$\hat{\beta}_3$ (x_3)	3.63	14.13	20.38	7.88	16.50	27.38
$\hat{\beta}_{12}$ (x_1x_2)				11.25	7.00	4.75
$\hat{\beta}_{23}$ (x_2x_3)	–19.00					
$\hat{\beta}_{11}$ (x_1^2)				–12.86	–5.79	–4.36
Error estándar	10.33	7.24	9.25	6.96	3.69	3.50

Método gráfico. En el primer paso se grafican y sobreponen las dos curvas de nivel de cada respuesta que corresponden a sus especificaciones. Se decide hacer los cortes en el sentido del plano (x_1, x_2), dejando a x_3 como el factor que determina los cortes. El resultado se observa en la figura 13.7 con los mejores cortes de dos regiones factibles, una en $x_3 = 0.81$ y otra en $x_3 = 0.70$. En el recuadro al interior de cada figura se reportan las otras dos coordenadas del punto óptimo de cada región, así como los valores predichos de las respuestas. Esta información se resume a continuación:

Solución	x_{10}	x_{20}	x_{30}	$\hat{Y}_1(x_0)$	$\hat{Y}_2(x_0)$	$\hat{Y}_3(x_0)$	$\hat{Y}_4(x_0)$	$\hat{Y}_5(x_0)$	$\hat{Y}_6(x_0)$
S1	1.0	0.85	0.70	186.0	175.2	173.6	191.5	172.0	183.5
S2	0.02	1.0	0.81	185.8	174.1	171.8	192.1	172.5	184.8

Observe que en este ejemplo no es necesario estrechar las restricciones, ya que con las especificaciones iniciales se obtienen regiones relativamente pequeñas. O sea, todo el método gráfico en este caso se reduce al primer paso. Sin embargo, es posible trabajar de modo más fino recorriendo en el segundo paso las especificaciones hasta 10% de su distancia al valor objetivo (véase ejercicio 11).

Método de la función de deseabilidad. Con la información de las especificaciones presentada arriba, las funciones de deseabilidad d_i se definen de acuerdo con la

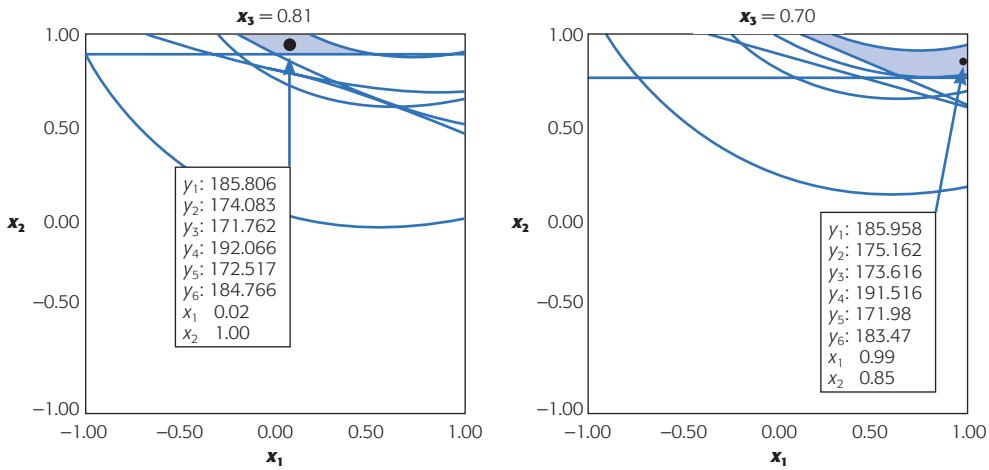


Figura 13.7 Soluciones con el método gráfico.

ecuación (13.2), usando exponentes y pesos unitarios. Al maximizar la DG se detectan dos soluciones que son muy similares a las del método gráfico. Las soluciones junto con su deseabilidad global son:

Solución	x_{10}	x_{20}	x_{30}	$\hat{Y}_1(x_0)$	$\hat{Y}_2(x_0)$	$\hat{Y}_3(x_0)$	$\hat{Y}_4(x_0)$	$\hat{Y}_5(x_0)$	$\hat{Y}_6(x_0)$	DG
S1	1.00	0.89	0.70	186.4	176.2	174.4	193.3	173.1	184.4	0.363
S2	0.06	1.00	0.80	185.9	174.3	171.9	192.9	172.9	185.0	0.289

Note que la variable más favorecida es Y_6, seguida por Y_4 y Y_1.

Trabajo con un software estadístico

El método de la función de deseabilidad para la optimización simultánea de varias respuestas se puede aplicar en varios software, entre ellos *Statgraphics*, *Minitab* y *Design Expert*. Este último cuenta con un graficador más avanzado. En cualquier software, el primer paso es ajustar un modelo jerárquico para cada variable y verificar su calidad.

Statgraphics

En este paquete una vez que se tiene un modelo ajustado para cada variable se aplica el método de la función de deseabilidad con la secuencia: *Special* → *Experimental design* → *Multiple response optimization*. En las versiones más recientes la secuencia es *DOE* → *Design Analysis* → *Multiple response optimization*. Después, se especifican las variables, que deben tener el modelo depurado en la sesión actual del software. Una vez que se entra al procedimiento, con el botón derecho del puntero (ratón) se abren las opciones de análisis (*Analysis options*) para definir las funciones de deseabilidad de cada respuesta. Luego, se declara lo que se quiere con cada Y (máximo, mínimo o cierto valor objetivo), así como los rangos deseables para cada una. Asimismo, se asignan los exponentes, valores entre 0.1 y 10 que dan forma a la función de deseabilidad dentro de los rangos declarados. El valor por omisión es 1,

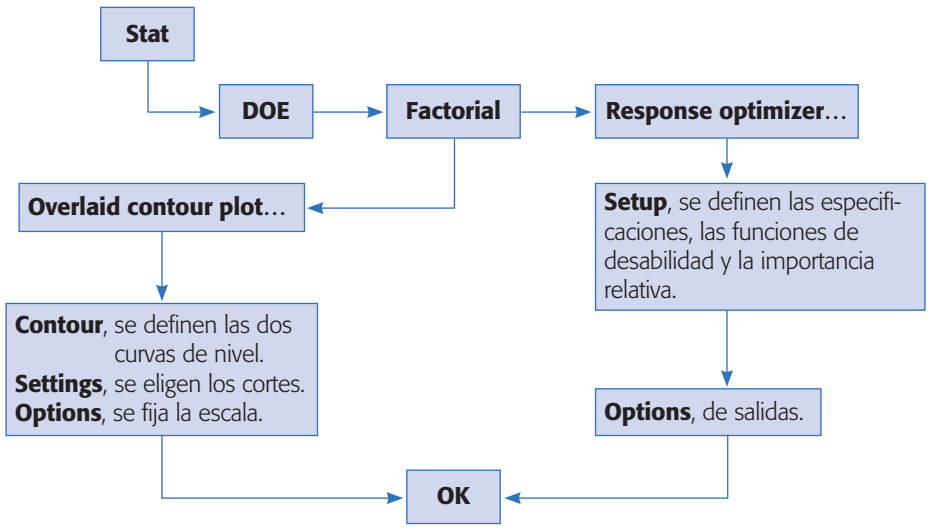


Figura 13.8 Esquema en *Minitab* para optimización simultánea.

que da lugar a deseabilidades lineales dentro del rango. También es preciso declarar la importancia relativa en una escala del 1 al 5. En la ventana *Optimize desirability* se reportan los resultados de la optimización.

Minitab

En este sistema se usa la secuencia de la figura 13.8, donde *Overlaid contour plot* es el graficador y *Response optimizer* es el método de la función de deseabilidad.

Design Expert

En el menú de *Optimization* de este software están las opciones *Graphical* (método gráfico) y *Numerical* (método de deseabilidad). Dentro de cada opción se entra a *Criteria* para definir toda la información sobre las variables, restricciones y pesos, según sea el caso. *Solutions* muestra los resultados numéricos y *Graphs* las figuras de contornos o de deseabilidad, según el método seleccionado.

Preguntas y ejercicios

1. ¿Por qué es necesario aplicar los métodos de optimización simultánea?
2. ¿Por qué se dice que los puntos óptimos simultáneos representan soluciones de compromiso?
3. Explique brevemente en qué consisten los métodos de optimización simultánea:
   a) Método gráfico.
   b) Método de la función de deseabilidad.
4. ¿Por qué la deseabilidad global se obtiene con la media geométrica de las deseabilidades individuales, y no con la deseabilidad promedio (media aritmética)?
5. ¿Qué información se necesita acerca de cada variable de respuesta para aplicar los métodos analizados en este capítulo?

6. ¿Cómo se trabaja con variables de respuesta con especificación unilateral?

7. Si en un experimento se quieren optimizar varias respuestas y una de ellas debe tomar valores de entre 50 y 70, pero en el experimento los resultados en los diferentes tratamientos para esas variables estuvieron entre 20 y 40, ¿cree que el método de optimización simultánea pueda encontrar un punto o región donde se alcancen los valores deseados para esa variable? En caso negativo, explique por qué y proponga alternativas.

8. Acerca del método gráfico de optimización simultánea:
   a) ¿Cuáles son sus ventajas?
   b) ¿Qué se debe hacer cuando no se observa la región factible en ningún corte?
   c) ¿A qué situación equivale lo anterior con el método de la función de deseabilidad?

9. ¿Cuál es la diferencia entre dar pesos a las variables y definir los exponentes de las funciones de deseabilidad?

10. Con respecto al problema del ejemplo 13.1 de optimización de neumáticos, realice lo siguiente:
    a) Ajuste un modelo de segundo orden jerárquico a cada respuesta y verifique que es adecuado.
    b) Dado que el modelo para Y_2 tiene un ajuste no muy adecuado, proponga transformaciones para esa variable por ejemplo del tipo log (Y_2) (véase capítulo 5), de tal forma que se mejore la calidad del modelo ajustado. En los siguientes incisos utilice este nuevo modelo.
    c) Determine la combinación de niveles de los factores que optimiza individualmente el valor de cada respuesta.
    d) Obtenga el punto óptimo simultáneo con el método gráfico. Utilice los criterios de reducción proporcional y de reducción proporcional al error estándar. Compare las soluciones a las que llegue.
    e) Construya las funciones de deseabilidad para cada variable y obtenga la función de deseabilidad global. Para definir las especificaciones para la variable Y_2 transformada, aplique la misma transformación a las especificaciones originales.
    f) Maximice la función de deseabilidad global y encuentre las condiciones óptimas simultáneas. Compare con la solución gráfica.
    g) ¿Cuáles son los valores esperados de las respuestas en las condiciones óptimas? ¿Cuál variable es la que se sacrifica más en aras de un mejor desempeño global?
    h) Compare con los resultados obtenidos en el ejemplo 13.1.

11. En el ejemplo 13.2 sobre semiconductores, rehaga el análisis gráfico considerando un estrechamiento de hasta 10% de la distancia entre las especificaciones y los valores nominales. Verifique que para una reducción mayor desaparecen las regiones factibles.

12. Algunas de las características importantes de los cubos de tomate son su textura y firmeza. Para proveer de buena textura y firmeza al producto final, el tomate, se procesa con sales de calcio. Se planea un experimento en el proceso de calcificación para estudiar cómo influyen los factores: concentración de calcio (x_1), temperatura de la solución (x_2) y tiempo de tratamiento (x_3) sobre las respuestas de interés que son: contenido total de calcio (Y_1), firmeza (Y_2) y pH (Y_3). Los factores con sus niveles de prueba, así como las variables de respuesta con sus rangos deseados se muestran en las siguientes tablas:

Variables de proceso	Var. codificada	−1	0	1
Concentración de calcio (% $CaCl_2$)	x_1	0.05	0.75	1.45
Temperatura de la solución (°C)	x_2	35	50	65
Tiempo de tratamiento (min)	x_3	0.5	2.0	3.5

Variables de respuesta	Rango óptimo	Valor deseado
Contenido total de calcio Ca, (mg/g)	700-800	menor que 800 ($Y_1 < 800$)
Firmeza (N/g)	20.0-20.5	mayor que 20 ($Y_2 > 20$)
pH	3.92-3.95	menor que 3.95 ($Y_3 < 3.95$)

Los datos observados en el experimento son:

Factores			Respuestas		
x_1	x_2	x_3	Y_1	Y_2	Y_3
1	1	0	3 720.0	23.04	3.86
1	–1	0	5 578.8	24.81	3.75
–1	1	0	390.4	15.40	4.35
–1	–1	0	248.1	12.85	4.20
1	0	1	7 490.2	26.48	3.61
1	0	–1	1 842.9	20.59	4.00
–1	0	1	253.5	14.81	4.33
–1	0	–1	152.0	10.10	4.35
0	1	1	2 890.3	20.79	4.01
0	1	–1	1 162.5	21.57	4.33
0	–1	1	1 698.1	22.85	3.69
0	–1	–1	804.9	17.75	3.92
0	0	0	1 505.9	23.53	3.85
0	0	0	1 274.3	20.00	4.13
0	0	0	1 660.3	24.12	3.77

a) Ajuste un modelo de segundo orden jerárquico para cada una de las variables de respuesta.
b) Verifique la calidad de los modelos y calcule el error estándar de cada uno de ellos.
c) Mediante el método gráfico encuentre una solución óptima simultánea para las tres respuestas.
d) Con el método de la función de deseabilidad encuentre una solución óptima simultánea.
e) Compare y comente las soluciones obtenidas con ambos métodos.

13. Box y Draper (1987) presentan un experimento sobre la calidad de un proceso de impresión de tinta a color sobre etiquetas. El valor objetivo de la respuesta promedio es 500, y se quiere lograr con variabilidad mínima. Los factores de control así como sus niveles en unidades originales son:

Factores/niveles	–1	0	1
x_1 : velocidad	30	45	60
x_2 : presión	90	110	130
x_3 : distancia	12	20	28

En la siguiente tabla se muestran los datos observados en tres réplicas del experimento, así como la media $\bar{Y}$ y la desviación estándar en cada tratamiento S.

x_1	x_2	x_3	Rep 1	Rep 2	Rep 3	$\bar{Y}$	S
–1	–1	–1	34	10	28	24	12.5
0	–1	–1	115	116	130	120.3	8.4
1	–1	–1	192	186	263	213.7	42.8
–1	0	–1	82	88	88	86.0	3.5
0	0	–1	44	178	188	136.7	80.4
1	0	–1	322	350	350	340.7	16.2
–1	1	–1	141	110	86	112.3	27.6
0	1	–1	259	251	259	256.3	4.6
1	1	–1	290	280	245	271.7	23.6
–1	–1	0	81	81	81	81.0	0.0
0	–1	0	90	122	93	101.7	17.7
1	–1	0	319	376	376	357.0	32.9
–1	0	0	180	180	154	171.3	15.0
0	0	0	372	372	372	372.0	0.0
1	0	0	541	568	396	501.7	92.5
–1	1	0	288	192	312	264.0	63.5
0	1	0	432	336	513	427.0	88.6
1	1	0	713	725	754	730.7	21.1
–1	–1	1	364	99	199	220.7	133.8
0	–1	1	232	221	266	239.7	23.5
1	–1	1	408	415	443	422.0	18.5
–1	0	1	182	233	182	199.0	29.4
0	0	1	507	515	434	485.3	44.6
1	0	1	846	535	640	673.7	158.2
–1	1	1	236	126	168	176.7	55.5
0	1	1	660	440	403	501.0	138.9
1	1	1	878	991	1 161	1 010.0	142.5

a) ¿Qué diseño se está empleando? ¿Es adecuado para ajustar un modelo de segundo orden?

b) Ajuste un modelo de segundo orden para $\bar{Y}$ y verifique que sea adecuado. ¿Qué porcentaje de la variabilidad explica?

c) Determine gráficamente un punto en la región experimental donde la respuesta promedio sea igual a 500. Encuentre un intervalo al 95% de confianza para la respuesta promedio en ese punto.

d) Ajuste un modelo de segundo orden para S, compruebe que es adecuado y encuentre el punto óptimo (minimizar) individual.

e) Repita el inciso anterior, pero ahora considerando la modelación de la transformación log(S), ¿cuál de los dos modelos es mejor?

f) Considere los mejores modelos ajustados para las respuestas $\bar{Y}$, S, haga un análisis de optimización simultánea si interesa un valor de 500 en la primera y un valor mínimo de la segunda respuesta. Defina las curvas de nivel y los rangos deseables para las dos respuestas, también aplique los métodos gráficos y de deseabilidad para proponer soluciones óptimas globales.

g) Exprese los puntos óptimos encontrados en las unidades originales del proceso.

14. Salcedo *et al*. (2002) realizan un experimento para evaluar y optimizar el efecto del pH de extracción (x_1) y del pH de precipitación (x_2) sobre tres características del proceso de obtención de aislados proteicos usando una variedad de semilla de amaranto. Las tres variables de respuesta que se quieren maximizar son: contenido de proteína (CP), índice de blancura (IB) y empatía de transición (ET). El experimento y los datos observados se muestran a continuación:

Trat.	x_1	x_2	CP	IB	TD
1	–1	–1	81.26	26.90	2.66
2	1	–1	83.35	25.24	2.38
3	–1	1	78.26	24.29	6.32
4	1	1	81.72	18.12	6.73
5	–1.41	0	78.08	29.34	4.12
6	1.41	0	81.90	21.17	4.07
7	0	–1.41	80.30	30.10	1.55
8	0	1.41	78.00	27.27	6.85
9	0	0	75.82	28.34	3.21
10	0	0	75.93	27.68	3.21

a) Obtenga modelos de regresión que expliquen el comportamiento de las respuestas.
b) Encuentre los puntos óptimos individuales de las tres respuestas, considerando la región experimental que va de –1.5 a 1.5 en cada factor codificado.
c) Haga el análisis gráfico simultáneo tomando como referencia para definir las restricciones de cada *Y* los valores mínimos y máximos observados en el experimento. Dibuje en una sola gráfica los óptimos individuales y el óptimo global.
d) Aplique el método de la función de deseabilidad considerando las mismas restricciones iniciales del método gráfico. Compare los resultados obtenidos.

15. Se corre un experimento central compuesto para optimizar de manera simultánea tres variables de densidad óptica (ODX4, ODX3 y ODX2.5), que se supone dependen de dos factores: iniciador (TAI) y material de color (CFM). Se busca maximizar las tres respuestas; 3, 3 y 2 son las cotas inferiores correspondientes y, desde el punto de vista práctico, se considera suficiente con lograr valores de 3.5, 3.5 y 2.5 para las respuestas. Los datos observados se muestran en la siguiente tabla:

TAI	CFM	ODX4	ODX3	ODX2.5
–1.0	–1.0	1.37	2.45	2.17
1.0	–1.0	2.29	3.04	2.38
–1.0	1.0	3.35	3.22	1.72
1.0	1.0	3.44	3.63	1.75
–1.41	0.0	2.79	2.87	1.89
1.41	0.0	2.80	3.38	2.04
0.0	–1.41	2.07	2.72	2.35
0.0	1.41	3.86	3.31	1.57
0.0	0.0	2.91	3.44	1.96
0.0	0.0	3.23	3.33	2.03
0.0	0.0	2.79	2.90	1.92
0.0	0.0	3.24	3.47	2.06
0.0	0.0	3.24	3.47	1.83

a) Aplique el método gráfico para encontrar un punto óptimo simultáneo para las tres respuestas; considere la región experimental de –1.5 a 1.5 para cada factor.
b) Dibuje el punto óptimo simultáneo, así como los puntos óptimos individuales sobre la misma figura.
c) Aplique el método de la función de deseabilidad para encontrar un punto óptimo global y compare con los resultados del inciso *a*).
d) Dibuje la función de deseabilidad utilizada para cada variable.
e) ¿Cuál es la diferencia entre los pesos (*weight*) y la importancia (*importance*) al aplicar el método de la función de deseabilidad?

Capítulo 14

Diseños anidados y diseños en parcelas divididas

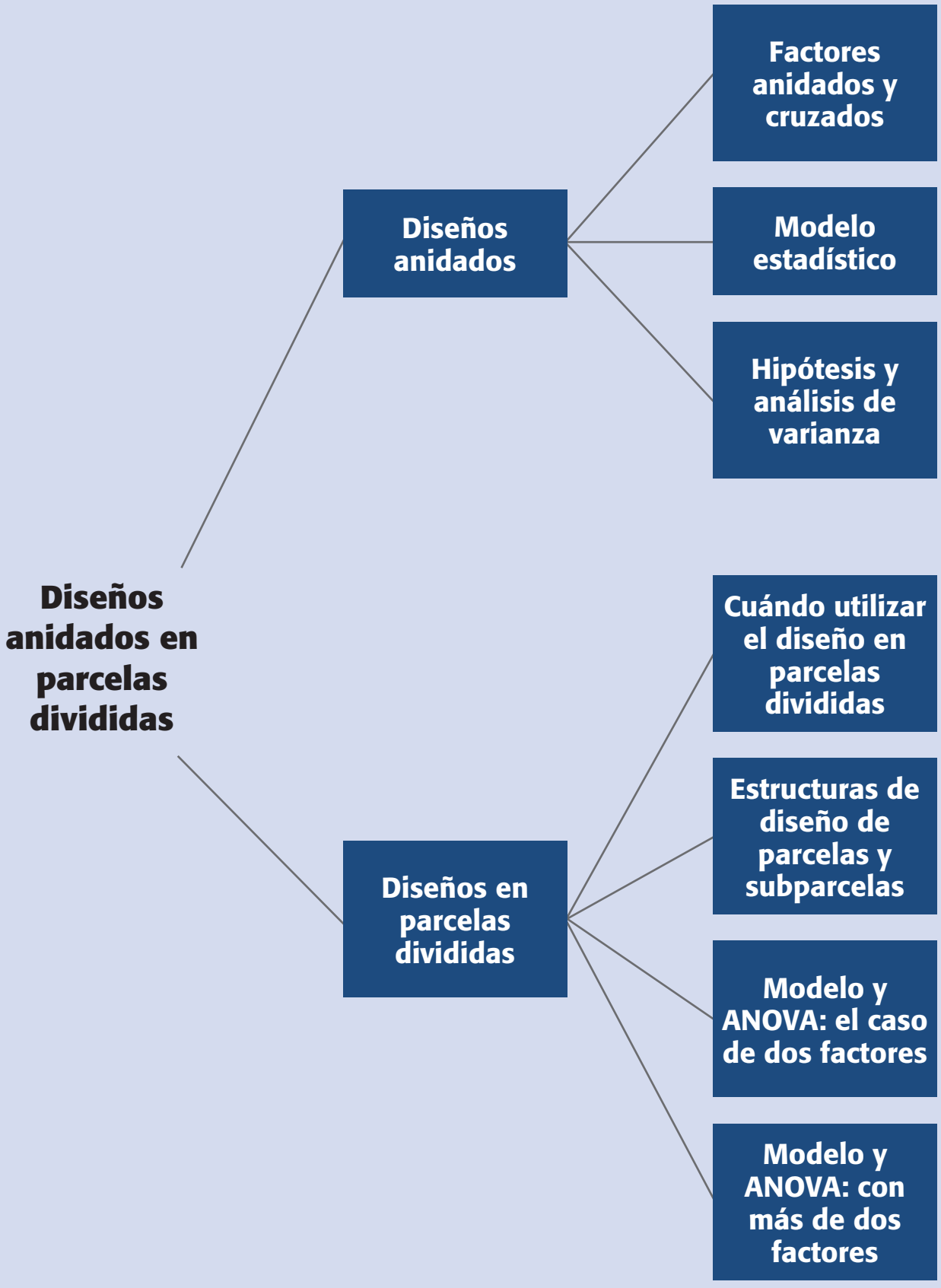
Sumario

- Diseños anidados
- Modelo y análisis estadístico del diseño anidado
- Diseños en parcelas divididas
- Modelo y análisis estadístico de los diseños en parcelas divididas
- Cómo hacer los cálculos usando software

Objetivos de aprendizaje

- Distinguir entre factores anidados y cruzados e identificar situaciones donde el diseño anidado es la opción experimental adecuada.
- Realizar el análisis de diseños anidados que involucran o no factores cruzados.
- Identificar diferentes situaciones prácticas que pueden llevar a utilizar un experimento en parcelas divididas.
- Distinguir diferentes estructuras de diseño que pueden presentar los factores en las parcelas y las subparcelas.
- Comprender por qué algunos experimentos con arreglo interno y externo son en realidad parcelas divididas.
- Analizar e interpretar correctamente el diseño en parcelas divididas. Entender de qué manera un análisis incorrecto podría cambiar las conclusiones.

Mapa conceptual


Diseños anidados en parcelas divididas
Diseños anidados
Factores anidados y cruzados
Modelo estadístico
Hipótesis y análisis de varianza
Diseños en parcelas divididas
Cuándo utilizar el diseño en parcelas divididas
Estructuras de diseño de parcelas y subparcelas
Modelo y ANOVA: el caso de dos factores
Modelo y ANOVA: con más de dos factores

Conceptos clave

- Arreglo interno-externo en parcelas divididas
- Diseño anidado o jerárquico
- Diseño en parcelas divididas
- Error de parcela
- Error de subparcela
- Estructura de diseño
- Factores cruzados
- Factorial aleatorio
- Factorial anidado
- Parcela
- Parcelas doblemente divididas
- Parcelas en bloques
- Subparcela
- Unidad experimental

Existen situaciones experimentales que involucran a varios factores y donde surgen consideraciones que se traducen en restricciones a la completa aleatorización de las corridas experimentales, o bien, hay situaciones en donde los niveles de prueba de un factor no son los mismos al cambiar los niveles de otro factor. Estos escenarios causan que el experimento deje de ser un factorial completamente aleatorizado para convertirse en un diseño en parcelas divididas (*Split-plot Design*) o en un diseño anidado (*Nested Design*). Recordemos que los diseños que usan bloques (capítulos 4 y 6) también surgen de restricciones a la aleatorización, y en este caso se incluyen los factores de bloques no porque sean de interés en sí, sino para incrementar la precisión del experimento. En cambio, en los diseños anidados y de parcelas divididas ésa no es la situación. Por ejemplo, en el caso de este último tipo de diseño, la restricción es impuesta por factores que son de interés para el experimentador, y se implementa para facilitar y hacer más eficiente la ejecución del experimento.

El error típico que se comete en las situaciones experimentales de los diseños anidados y de los diseños en parcelas divididas es que, a menudo, el experimentador los analiza como si fueran diseños factoriales, y esto puede llevar a conclusiones erróneas, como lo ilustraremos a través de ejemplos.

Diseños anidados

Factores cruzados
Sus niveles se pueden combinar aleatoriamente sin ninguna restricción en cada nivel del otro factor.

En capítulos previos vimos que cuando se tiene un diseño factorial con dos **factores cruzados** *A* y *B*, se corren en orden aleatorio todas las posibles combinaciones de niveles de los dos factores. Ahí, los niveles de cada factor se pueden combinar en cualquier momento con los niveles del otro factor, y en este caso los niveles de un factor son exactamente los mismos que en cada nivel del otro factor. Por otra parte, cuando se dice que el factor *B* está anidado en el factor *A* significa que los niveles del factor *B* no son los mismos en cada nivel del factor *A*. Es decir, hay una especie de relación padre-hijo entre los niveles del factor *A* (padres) y los niveles del factor *B* (hijos). En este sentido, los **diseños anidados** también se conocen como *diseños jerárquicos*.

Diseño anidado o jerárquico
Se aplica cuando los niveles de un factor *B* están anidados en los niveles de otro factor *A* de mayor jerarquía. Al factor *B* se le llama factor anidado y sus niveles son diferentes en cada nivel de *A*.

Ejemplo 14.1

Un caso típico de anidamiento surge cuando se quiere evaluar el error de un sistema de medición mediante un estudio de repetibilidad y reproducibilidad (R&R), ver Gutiérrez y De la Vara (2004), donde la medición de cada pieza se obtiene por medio de una prueba destructiva. Los estudios R&R requieren que una misma pieza sea medida por lo menos dos veces por los operadores participantes en el estudio, pero esto no es posible en pruebas destructivas. Si suponemos que cada pieza o parte se puede dividir en tres pedazos que se destruyen al medirlos, entonces estos tres pedazos sólo permiten que un operador haga sus repeticiones sobre la misma pieza; por ello, los otros operadores tendrán que utilizar piezas diferentes. En este sentido se dice que las piezas están anidadas en los operadores, como se muestra en la figura 14.1, donde a fin de ilustrar el diseño anidado se consideran: dos operadores, cuatro piezas y tres repeticiones.

Advierta que en la figura 14.1 se enfatiza que cada operador utiliza piezas diferentes empleando la numeración de las piezas del 1 al 8. Si cada pieza no se pudiera

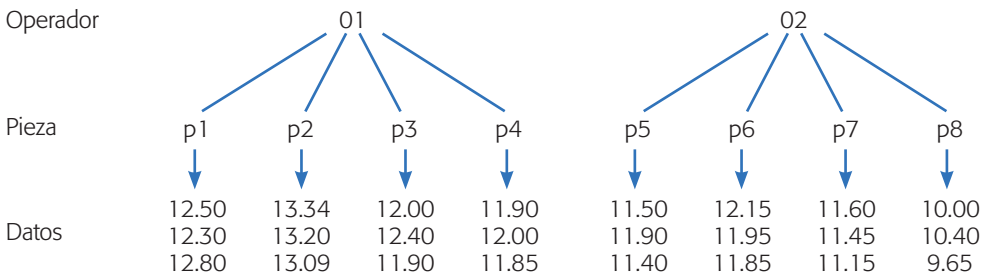


Figura 14.1 Estudio R&R anidado, con dos operadores, cuatro piezas y tres repeticiones por pieza.

dividir en tres pedazos y/o la pieza sólo resistiera una medición, todavía se podría correr este diseño, pero suponiendo que es posible construir pequeños lotes de tres piezas similares, donde cada operador haría sus tres repeticiones. En este experimento es razonable suponer que el factor operador es fijo porque los dos trabajadores están disponibles y el factor pieza es aleatorio (ver capítulo 5), ya que éstas son una selección aleatoria de la producción.

Un aspecto que distingue al diseño anidado es que no tiene sentido considerar la interacción entre los factores con respecto al anidamiento, ya que por construcción, los niveles del factor anidado son diferentes en cada nivel del factor principal. Para que la interacción tenga sentido los operadores deben medir exactamente las mismas piezas, como ocurre en un diseño factorial cruzado.

En este primer ejemplo se tienen sólo dos factores, uno anidado en el otro, pero existen situaciones más complejas, por ejemplo de tres factores, con C anidado en B, y B anidado en A. También hay situaciones experimentales en donde algunos factores pueden cruzarse o combinarse de acuerdo a un arreglo factorial mientras otros factores están anidados. Esto se ilustra en el siguiente ejemplo.

Ejemplo 14.2

Factores cruzados y anidados. Consideremos el caso de un ingeniero industrial que quiere comparar dos métodos de ensamble (factor A), utilizando dos arreglos diferentes del área de trabajo (factor B) y cuatro operadores (factor C). Como las áreas de trabajo donde se implementa cada arreglo están físicamente separadas, se deben utilizar cuatro operadores diferentes en cada arreglo. En otras palabras, los operadores están anidados en los arreglos. Mientras que en los métodos de ensamble están cruzados con los arreglos y con los operadores. En este diseño, conocido como *factorial-anidado*, los datos se pueden acomodar como en la figura 14.2, la variable de respuesta es el tiempo de ensamble.

Diseño factorial-anidado
Diseño que involucra factores cruzados y anidados.

Modelo y análisis estadístico del diseño anidado

En el diseño anidado más simple (anidamiento en dos etapas), que es el caso del ejemplo 14.1, se tienen dos factores A y B, con los niveles de B anidados en cada

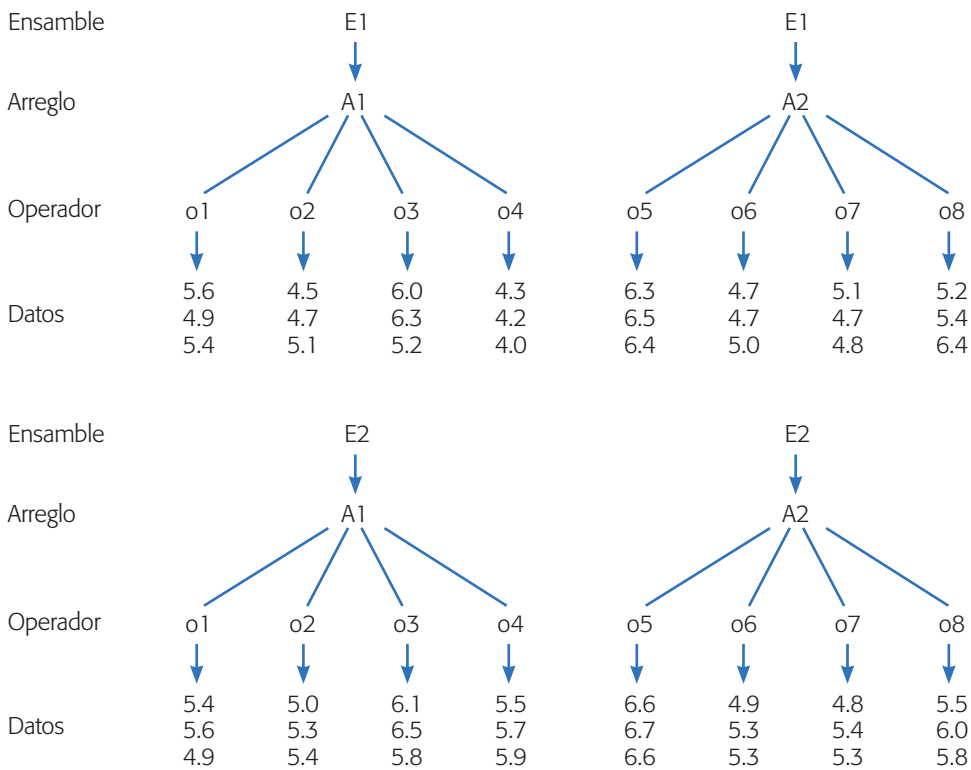


Figura 14.2 Comparación de métodos de ensamble, diseño factorial-anidado.

nivel de A. El modelo estadístico que describe el comportamiento esperado de la respuesta en este diseño está dado por:

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{k(ij)}; \quad i = 1, 2, ..., a; \; j = 1, 2, ..., b; \; k = 1, 2, ..., n \quad \textbf{(14.1)}$$

donde hay a niveles de A, b niveles del factor B anidados bajo cada nivel de A y n réplicas. Además μ es la media general, α_i es el efecto del i-ésimo nivel del factor A, $\beta_{j(i)}$ es el efecto del j-ésimo nivel del factor B, que está anidado en el nivel i del factor A, y $\varepsilon_{(ij)k}$ es el error aleatorio con los supuestos usuales de normalidad, varianza constante e independencia. Note que el modelo no incluye el concepto de interacción, ya que el factor B no está cruzado sino anidado en el factor A. Las hipótesis de interés expresadas en la notación del modelo son:

$$\begin{gathered}
H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0 \\
H_A : \alpha_i \neq 0 \text{ para algún } i, \\
H_0 : \beta_{1(1)} = \beta_{2(1)} = \cdots = \beta_{b(1)} = 0 \\
H_A : \beta_{j(1)} \neq 0 \text{ para algún } j, \\
\vdots \\
H_0 : \beta_{1(a)} = \beta_{2(a)} = \cdots = \beta_{b(a)} = 0 \\
H_A : \beta_{j(a)} \neq 0 \text{ para algún } j
\end{gathered} \quad \textbf{(14.2)}$$

Observe que las hipótesis sobre el factor anidado se expresan dentro de cada nivel del factor A usando la notación $j(i)$, que significa nivel j anidado en el nivel i. Estas hipótesis se prueban mediante un ANOVA apropiado, el cual toma en cuenta si los factores son fijos o aleatorios. La mayoría de las sumas de cuadrados que se deben calcular para el ANOVA son las mismas que se calculan en el diseño factorial con dos factores (véase capítulo 6), salvo la suma de cuadrados del efecto anidado $B(A)$ que se denota por $SC_{B(A)}$ y que se obtiene con respecto a la media dentro de cada nivel del factor A y no con relación a la media global como se hace con el efecto A no anidado. Específicamente, la suma de cuadrados totales está dada por:

$$SC_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk}^2 - \frac{Y_{...}^2}{N} \qquad \textbf{(14.3)}$$

donde $N = abn$ es el total de observaciones en el experimento. Las sumas de cuadrados de los efectos A y $B(A)$ son:

$$SC_A = \sum_{i=1}^{a} \frac{Y_{i..}^2}{bn} - \frac{Y_{...}^2}{N} \qquad \textbf{(14.4)}$$

$$SC_{B(A)} = \sum_{i=1}^{a}\left(\sum_{j=1}^{b} \frac{Y_{ij.}^2}{n} - \frac{Y_{i..}^2}{bn} \right) \qquad \textbf{(14.5)}$$

y la suma de cuadrados del error, por diferencia, es:

$$SC_E = SC_T - SC_A - SC_{B(A)}$$

donde los grados de libertad correspondientes cumplen esta misma relación,

$$ab(n-1) = (abn-1) - (a-1) - a(b-1)$$

Al igual que en los diseños factoriales, los valores esperados de los cuadrados medios en un experimento anidado dependen de si ambos factores se consideran *fijos* o *aleatorios*; o uno fijo y el otro aleatorio (véase capítulo 5). En la práctica se suele abusar del supuesto de factores fijos, quizá porque los paquetes estadísticos tienden a suponer lo mismo, o porque es más fácil comprender las hipótesis de interés expresadas en términos de los efectos promedio, como se muestran en las ecuaciones (14.2). Cabe recordar que cuando los efectos son aleatorios, las hipótesis de interés se expresan en términos de las varianzas de los mismos (H_0: $\sigma^2_{efecto} = 0$). En resumen, para el caso anidado con dos factores, en la tabla 14.1 se muestran los valores esperados de los cuadrados medios para las tres combinaciones de interés de factores fijos y aleatorios. La combinación A aleatorio y B fijo usualmente no es de interés práctico, por eso no se incluye.

Factor aleatorio
Los niveles de prueba utilizados son una muestra aleatoria de la población de niveles posibles para este factor.

La información que provee esta tabla es útil para saber, en cada caso, cómo se deben construir los estadísticos de prueba para las hipótesis de interés y, por lo tanto, para saber cómo se calculan las dos últimas columnas de la tabla de ANOVA. Por ejemplo, en el caso (fijo, fijo) los estadísticos se construyen con los cocientes,

$$F_0^A = \frac{CM_A}{CM_E} \qquad \text{y} \qquad F_0^{B(A)} = \frac{CM_{B(A)}}{CM_E} \qquad \textbf{(14.6)}$$

Tabla 14.1 Valores esperados de los *CM* en el diseño anidado con dos factores.

E(CM)	*A* fijo *B* fijo	*A* fijo *B* aleatorio	*A* aleatorio *B* aleatorio
$E(CM_A)$	$\sigma^2 + \frac{bn\sum \alpha_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + \frac{bn\sum \alpha_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + bn\sigma_\alpha^2$
$E(CM_{B(A)})$	$\sigma^2 + \frac{n\sum\sum \beta_{j(i)}^2}{a(b-1)}$	$\sigma^2 + n\sigma_\beta^2$	$\sigma^2 + n\sigma_\beta^2$
$E(CM_E)$	σ^2	σ^2	σ^2

que serían igual a uno bajo las respectivas hipótesis nulas; en la tabla 14.1 observe que cuando los efectos A y $B(A)$ son nulos o iguales a cero, los estadísticos de (14.6) se reducirían a $\sigma^2/\sigma^2 = 1$. De manera similar, de la columna (A fijo, B aleatorio) se ve que para esta combinación los estadísticos correspondientes son:

$$F_0^A = \frac{CM_A}{CM_{B(A)}} \quad \text{y} \quad F_0^{B(A)} = \frac{CM_{B(A)}}{CM_E} \tag{14.7}$$

Observe que en los estadísticos de prueba (ecuaciones 14.6 y 14.7) cambia el denominador, dependiendo de lo que se supone con respecto a los factores. La información hasta aquí presentada es fácil de bosquejar en la tabla de ANOVA para cada caso. Veámoslo, pero en el contexto de un ejemplo.

Ejemplo 14.1 (Continuación)

Con los datos de la figura 14.1 para el estudio R&R y suponiendo el factor operador (A) fijo y pieza (B) aleatorio, el ANOVA anidado se muestra en la tabla 14.2, en donde se observa un efecto contundente de las piezas, que es lo esperado en este tipo de estudios porque indica que el sistema de medición es capaz de distinguir las piezas. Algo indeseable en los estudios R&R es el efecto significativo del operador, que en este caso, dado que su valor-p es 0.0606, y está muy cerca de 0.05, su impacto en la medición puede ser relevante.

Debido a que este ejemplo es un caso de estudios R&R, el error representa la repetibilidad, la fuente pieza (operador) es la variación de las piezas y la fuente operador representa la reproducibilidad (véase Gutiérrez y De la Vara (2004), capítulo 11). Los valores estimados de cada componente de varianza se pueden obtener al despejar los cuadrados medios correspondientes de la tabla 14.1. En el ejemplo como A es fijo y B aleatorio, los componentes de varianza son estimados por:

$$\hat{\sigma}^2_{error} = CM_{error};\ \hat{\sigma}^2_{B(A)} = \frac{CM_{B(A)} - CM_E}{n} \quad \text{y} \quad \hat{\sigma}^2_A = \frac{CM_A - CM_{B(A)}}{bn} \tag{14.8}$$

y con ellos se calcula la variación que se atribuye al sistema de medición, como $\hat{\sigma}_{R\&R} = \sqrt{\hat{\sigma}^2_{emor} + \hat{\sigma}^2_A}$. Ésta se expande y compara con la tolerancia $[5.15\hat{\sigma}_{R\&R}/(ES - EI)]$, y con la variación total $(5.15\hat{\sigma}_{R\&R}/5.15\hat{\sigma}_{total})$. El sistema de medición no tiene la

Tabla 14.2 ANOVA anidado. Ejemplo 14.1.

FV	SC	GL	CM	F_0	Valor-*p*
Operador	8.4966	1	8.4966	5.3187	0.0606
Pieza(Operador)	9.5849	6	1.5975	28.9444	0.0000
Error	0.8831	16	0.0552		
Total	18.9646	23			

precisión adecuada (mucho error) cuando cualquiera de estos índices es mayor que 0.3 (30%), pero el sistema es bueno si ambos son menores que 20% (véase Gutiérrez y De la Vara (2004), capítulo 11).

En los software estadísticos usualmente se incluyen los diseños anidados, por lo que no es complicado obtener el ANOVA correspondiente. En caso contrario, se puede obtener fácilmente de la tabla de ANOVA del diseño factorial usual (no anidado), simplemente se debe observar que la suma de cuadrados anidada $SC_{B(A)}$ es igual a la suma $SC_B + SC_{AB}$ del diseño factorial usual. En este ejemplo se puede ver que:

$$SC_{pieza(operador)} = SC_{pieza} + SC_{operador \times pieza} \qquad \textbf{(14.9)}$$

lo cual se deja como ejercicio al lector (ejercicio 7).

Ejemplo 14.2 (Continuación)

Retomando el problema del ejemplo 14.2 y los datos de la figura 14.2, donde se involucran tres factores: el factor *A* (método de ensamble), cruzado con *B* (tipo de arreglo) y con *C* (operador que hace el ensamble). Los niveles del factor *C* están anidados en los niveles de *B*. El modelo estadístico para este tipo de diseño está dado por:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{k(j)} + (\alpha\gamma)_{ik(j)} + \varepsilon_{l(ijk)}; \quad i = 1, 2;\ j = 1, 2;\ k = 1, 2, \ldots, 4;\ l = 1, 2, 3 \qquad \textbf{(14.10)}$$

donde μ es la media general, α_i es el efecto del ensamble i, β_j es el efecto del arreglo j, $(\alpha\beta)_{ij}$ es la interacción entre ensamble y arreglo, $\gamma_{k(j)}$ es el efecto del operador k anidado en el j-ésimo arreglo, $(\alpha\gamma)_{ik(j)}$ es el efecto de interacción ensamble con operador anidado en el j-ésimo arreglo y $\varepsilon_{l(ijk)}$ es el error aleatorio asociado con la observación l-ésima del tratamiento ijk. En la tabla 14.3 se muestra la tabla de ANOVA de este modelo declarando el factor operador aleatorio. Es muy importante no olvidarse de especificar si los factores son aleatorios o fijos, porque eso puede cambiar de manera radical las conclusiones (véase ejercicio 8). En esta tabla se observa que son significativos los efectos operador (arreglo), EnsaXOper (arreglo) y prácticamente ensamble (dado el nivel tan pequeño del valor-*p*).

La razón *F* se obtiene considerando al factor operador aleatorio, con lo que la mayoría de los cuadrados medios se divide entre el *CM* del efecto anidado de menor jerarquía: *AC*(*B*), con excepción de *A*:*Arreg* que se divide con el *CM* de *C*(*B*) y el propio *AC*(*B*) que se contrasta con el *CM* del error. Por supuesto que es más fácil usar un software con una opción de ANOVA que considere factores anidados y cruzados.

Tabla 14.3 ANOVA para tiempos de ensamble. Ejemplo 14.2.

FV	SC	GL	CM	F	Valor-*p*
A:Ensa	2.0419	1	2.0419	5.94	0.051
B:Arreg	0.7752	1	0.7752	0.34	0.579
AB	0.3169	1	0.3169	0.92	0.374
C(*B*):Oper(Arreg)	13.5529	6	2.2588	6.57	0.019
AC(*B*)	2.0629	6	0.3438	3.38	0.011
Error	3.2533	32	0.1017		
Total	22.0031	47			

Para interpretar los efectos significativos se utilizan las gráficas de efectos usuales, pero dentro de cada nivel del factor de mayor jerarquía. En la figura 14.3 se representa el efecto de interacción *AC*(*B*) dentro de cada nivel de *B* (cada arreglo). De la figura se concluye que el mejor método de ensamble es el método 1 (línea sólida), que se mantiene siempre por debajo. El efecto individual del arreglo que es no significativo en el ANOVA, sería la diferencia en los tiempos promedios de los dos gráficos. Mientras que el efecto de *C*(*B*), que sería el promedio de las líneas punteada y sólida en cada gráfica, resulta muy diferente en cada arreglo, en principio porque se consideran diferentes personas en cada lugar de trabajo. En conclusión, habría que entrenar más a los operadores para homogenizar y bajar los tiempos del método de ensamble 1 (véase figura 14.3), mediante cualquiera de los dos arreglos.

Si no se cuenta con un software estadístico que obtenga directamente el ANOVA de la tabla 14.3, éste se puede obtener a partir de la tabla de ANOVA del análisis factorial que ignora el anidamiento, pero considerando que los términos anidados son sumas de dos términos factoriales, específicamente,

$$SC_{Oper\ (Arreg)} = SC_{Oper} + SC_{Oper \times Arreg}$$
$$SC_{Ensa \times Oper\ (Arreg)} = SC_{Ensa \times Oper} + SC_{Ensa \times Oper \times Arreg} \qquad \textbf{(14.11)}$$

Parcela
Son unidades experimentales grandes a las cuales se asignan los niveles de factores difíciles de mover. En agricultura es un terreno para siembra.

Subparcela
Son unidades experimentales pequeñas que resultan de dividir la parcela. En las subparcelas se asignan los niveles de factores fáciles de manipular.

Diseño en parcelas divididas

El *diseño en parcelas divididas* (*split-plot*), cuyo nombre se deriva de su origen en agricultura, surge por una restricción particular para aleatorizar el orden de corrida de los tratamientos. Por ejemplo, suponga que se quieren comparar varios fertilizantes, cada uno de ellos se aplica a una parcela completa o grande, pero dicha parcela a su vez se divide en parcelas pequeñas o subparcelas para estudiar diferentes variedades de un cultivo. De esta manera, se tiene el factor de parcela (fertilizante) y el factor de subparcela (variedad). El hecho relevante es que los fertilizantes se aleatorizan al asignarlos a las *parcelas* y las variedades al asignarlas a las *subparcelas* (pero dentro de cada parcela); por ello, no se tiene, ni es conveniente una aleatorización completa de fertilizantes y variedades, como se haría con un diseño factorial. Una representación gráfica de las parcelas y subparcelas en este experi-

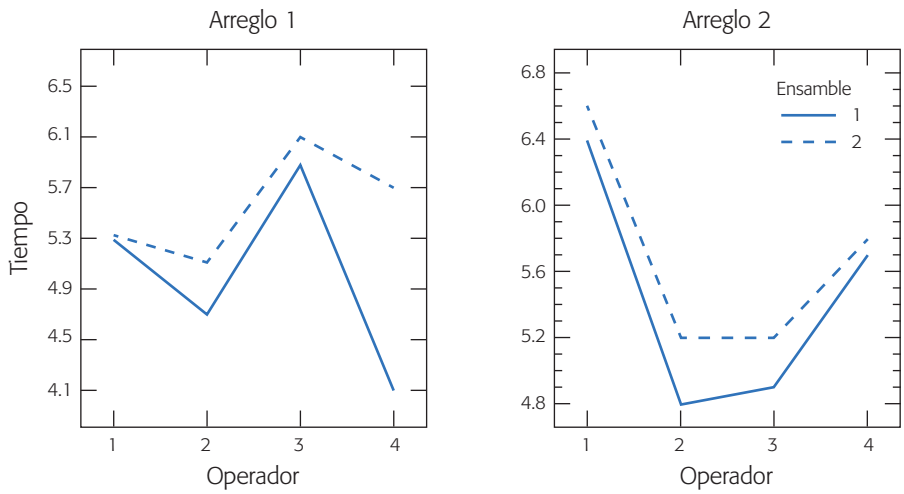


Figura 14.3 Interacción *AC*(*B*): Ensa×Oper (Arreg), ejemplo 14.2.

mento agrícola, considerando la asignación aleatoria de tres fertilizantes ($F1$, $F2$ y $F3$), cuatro variedades ($V1$, $V2$, $V3$ y $V4$) en tres réplicas (bloques I, II, III) se muestra en la figura 14.4. Así cada columna es una parcela.

Situaciones similares a la descrita se presentan en otras áreas experimentales, donde las parcelas y subparcelas ya no son físicamente pedazos de tierra, pero la idea es la misma: habrá tratamientos que se tienen que aplicar de manera aleatoria a *unidades experimentales* grandes (parcelas), y habrá otros tratamientos que se pueden aplicar al azar en unidades experimentales chicas (subparcelas), que resultan al dividir las primeras. Así, los primeros tratamientos son combinaciones de niveles de los factores en las parcelas y los segundos son combinaciones de niveles de los factores en las subparcelas. A menudo, los tratamientos en la parcela se aplican antes que los tratamientos en la subparcela. Estos diferentes tamaños de unidades experimentales provocan la aparición de dos estructuras de error que deben tomarse en cuenta al momento del análisis estadístico.

Unidad experimental
Individuo, objeto o cantidad de material a la que se le aplica un tratamiento para obtener un dato de cada respuesta de interés.

Cuándo utilizar el diseño de parcelas divididas

El usuario típico no suele distinguir cuándo y cómo puede utilizar el *diseño en parcelas divididas* y aprovechar sus ventajas, o cuando menos hacer el análisis correcto de una situación experimental que cae en este tipo de diseño. Existen diferentes motivaciones para considerar el diseño en parcelas divididas como el más apropiado. Por ejemplo, algunas situaciones específicas donde es recomendable este tipo de diseño son las siguientes:

Diseño en parcelas divididas
Experimento que involucra diferentes tamaños de unidades experimentales, que restringen la completa aleatorización de los tratamientos. Cada tipo de unidad genera un error experimental que es necesario estimar.

1. Cuando en un estudio experimental se tienen factores que son difíciles de cambiar de un nivel a otro, y también se tienen factores que son fáciles de mover. En general, los factores difíciles de manipular dificultan el trabajo experimental en un diseño factorial, ya que se tienen que correr en forma aleatoria los tratamientos, lo cual obliga a cambiar con frecuencia los niveles de tales factores. Una posible solución es la que se expuso en el ejemplo

F2	F1	F3	F1	F2	F3	F2	F3	F1
V1	V3	V4	V2	V2	V1	V3	V4	V4
V3	V4	V2	V4	V1	V3	V2	V1	V1
V4	V1	V3	V1	V3	V2	V4	V3	V4
V2	V2	V1	V3	V4	V4	V1	V1	V3
	Bloque I			Bloque II			Bloque III	

Figura 14.4 Diseño en parcelas divididas para tres fertilizantes (parcelas) y cuatro variedades (subparcelas) en tres réplicas (bloques).

10.1. La otra, y más efectiva es aplicar un diseño en parcelas divididas, donde los factores difíciles de mover se asignen a las parcelas y los fáciles de mover a las subparcelas, minimizando de esta manera los cambios de nivel de los factores difíciles durante el experimento. Un ejemplo típico de esta situación se da en los procesos de horneado (de pasteles, de metal, etc.) donde se dificulta estar cambiando el nivel de temperatura del horno; una mejor estrategia es elegir al azar uno de los niveles de temperatura con los que se desea experimentar y correr en ella todas las combinaciones de los factores restantes. De esta manera, el factor temperatura define las parcelas grandes y el resto de los factores serían las parcelas chicas.

2. Por su naturaleza, algunos factores requieren que la unidad experimental sea de gran tamaño, mientras que otros factores se pueden manejar a menor escala. Este caso se mostró en la figura 14.3, donde se quieren comparar varios fertilizantes pero es técnicamente impráctico aplicarlos a parcelas pequeñas, es decir, cada fertilizante sólo se puede aplicar en parcelas grandes. Éstas se dividen en subparcelas para estudiar diferentes combinaciones de otros factores, como variedades, prácticas de manejo, etc. Otro ejemplo de un factor de parcela grande es el fumigante, suponiendo que éste se tiene que aplicar desde una avioneta que no puede maniobrar en una parcela chica sin contaminar a la parcela de al lado.
3. En ocasiones se requiere incluir un factor adicional en aras de un mayor alcance de las conclusiones del experimento, pero la importancia de este factor adicional es menor. Por ejemplo, supongamos que se quieren comparar cuatro medicamentos para una enfermedad, y además controlar la dosis y el tiempo de aplicación en dos niveles cada uno. Pero si adicionalmente se considera que en estudios de este tipo la hora del día en la que se aplica el tratamiento puede influir en los resultados, entonces se incluye la hora en el experimento, dividiendo cada día en cuatro parcelas. Así, se aleatoriza a cuál de las cuatro parcelas (hora de aplicación) se asigna cada uno de los cuatro medicamentos. Dentro de cada parcela se aleatorizan las cuatro combinaciones de dosis y tiempo (subparcela) que corresponden a cada medicamento. Y esto se puede repetir varios días (bloque).
4. Otro caso es cuando se espera de antemano, por conocimiento previo, que algunos factores tengan efectos grandes sobre la respuesta, mientras que para otros factores se espera un efecto pequeño. En estas circunstancias, los niveles de los factores con efectos grandes se asignan de manera aleatoria a

las parcelas y los niveles de los factores con efectos pequeños se asignan aleatoriamente a las subparcelas. En un diseño en parcelas divididas los efectos de los factores en las parcelas se estiman con menor precisión que los efectos de los factores en las subparcelas, ya que existe más variabilidad entre las primeras dado su mayor tamaño. De aquí que el factor con efecto grande se asigne a la parcela grande, puesto que no se requiere mucha precisión para detectar un efecto de este tipo.

5. Muchos experimentos robustos que usan el diseño con arreglo interno y externo propuesto por Taguchi (capítulo 9) son, por la manera en que fueron realizados, diseños en parcelas divididas. Típicamente, cada combinación de los factores de proceso en el arreglo interno es una parcela, y cada combinación de los factores de ruido en el arreglo externo es una subparcela.

Para complementar lo anterior, en un diseño en parcelas divididas es de suma importancia decidir cuáles factores se irán a la parcela y cuáles a la subparcela, ya que de esto depende el tamaño de la parcela que será usada con cada factor y la precisión con que se estimarán los efectos. Todo esto sin olvidar las prácticas de manejo y facilidad para mover niveles, que son criterios a tomar en cuenta. Otras consideraciones que deben asignarse a las subparcelas son: si un factor es muy importante, sus niveles se mueven fácilmente o requiere poca cantidad de material experimental.

Ejemplo 14.3

Caso con dos factores. Potcner y Kowalski (2004) describen un experimento que tiene que ver con la propiedad de resistencia de la madera al agua. En el experimento se estudian dos tipos de pretratamientos y cuatro marcas de barniz. La variable de respuesta es la resistencia. El experimento se realizó de la siguiente manera: primero se selecciona al azar un pretratamiento y se aplica a un panel de madera. Después, el panel se parte en cuatro pedazos y a cada trozo se aplica una marca de barniz elegida al azar. Luego se aplica el segundo pretratamiento a otro panel, que se divide a su vez en cuatro pedazos donde se aplican los barnices al azar. Esto se hace tres veces. La resistencia obtenida en cada combinación de pretratamiento y barniz se muestra en la tabla 14.4. Por la manera en que se corrió el experimento, cada panel corresponde a una parcela y cada trozo de panel es una subparcela. De modo que en total se tienen seis parcelas y 24 subparcelas. Cada renglón son los datos observados en una parcela y cada casilla representa el dato en una subparcela.

Si no se supiera cómo se obtuvieron los datos, éstos se podrían tomar como un diseño en bloques completos al azar (capítulo 4), o como un diseño factorial 4×2 (capítulo 5) con tres réplicas, pero de acuerdo con el procedimiento experimental descrito, se trata de un diseño en parcelas divididas. Existen varios aspectos característicos del diseño en parcelas divididas que lo delatan como tal: los factores tienen dos tamaños de unidades experimentales; el panel completo para los pretratamientos y los pedazos de panel para los barnices. El procedimiento de aleatorización consta de dos pasos. Además, el número de repeticiones no es el mismo para cada factor, ya que pretratamiento tiene tres repeticiones y barniz seis. Más adelante se analizan los datos de la tabla 14.4.

Los diseños en parcelas divididas presentan complicaciones dependiendo de la situación experimental. Los problemas más comunes ocurren cuando:

Tabla 14.4 Datos de resistencia de la madera al agua, ejemplo 14.3.

	Barnices				
Pretratamiento	*B*1	*B*2	*B*3	*B*4	Media
1	43.0	51.8	40.8	45.5	45.28
1	57.4	60.9	51.1	55.3	56.18
1	52.8	59.2	51.7	55.3	54.75
2	46.6	53.5	35.4	32.5	42.00
2	52.2	48.3	45.9	44.6	47.75
2	32.1	34.4	32.2	30.1	32.20

Estructura de diseño
Maneras de acomodar los factores de parcela o de subparcela, que se reflejan en el modelo y su análisis. Puede ser completamente al azar, en bloques, factorial, etcétera.

1. Los niveles del factor parcela están metidos en una *estructura de diseño* completamente al azar (como en el ejemplo 14.3), de bloques completos al azar o de cuadro latino. Mientras que los niveles del factor subparcela también tienen diferentes arreglos dentro de cada parcela.
2. Existen más de dos factores tanto a nivel de parcela como a nivel de subparcela, como ocurre en muchos diseños con arreglo interno y externo propuestos por Taguchi (capítulo 9 y ejemplo 14.4). Cuando son muchos los factores en la subparcela es necesario utilizar factoriales fraccionados, pero la estructura de confusión resultante es diferente a la fracción por tratarse de un diseño en parcelas divididas (Bingham *et al*., 2004).
3. La subparcela se parte en subsubparcelas para dar lugar a otro factor de interés, en cuyo caso el experimento se llama *diseño en parcelas doblemente dividid*as (*Split-split-plot design*).

Parcelas doblemente divididas
Cuando a su vez la subparcela se divide para dar lugar a subsubparcelas donde se asignan los niveles de otro factor de interés.

Ejemplo 14.4

Caso con varios factores. Consideremos un ejemplo del tipo diseño robusto con arreglos interno y externo (véase capítulo 9), donde el objetivo del experimento es diseñar una harina preparada para pastel que sea robusta a las variaciones de temperatura y tiempo de preparación por parte de los consumidores. Para ello se plantea el experimento de la tabla 14.5, donde se tienen tres factores de diseño (control): harina (*H*), grasa (*G*) y polvo de huevo (*P*) acomodados en un factorial 2^3. Los factores de ruido son temperatura del horno (*T*) y tiempo de horneado (*Ti*). La variable de respuesta es el resultado de la evaluación promedio de un panel de expertos en una escala de 1 a 7, donde siete es la máxima evaluación positiva. Para sintetizar, nos referimos a la variable como el sabor.

El experimento de la tabla 14.5 es un diseño en parcelas divididas porque se corrió de la siguiente manera: se amasó una cantidad suficiente de harina en cada combinación de los factores de diseño, y luego ésta se dividió en cuatro lotes con los cuales se prepararon cuatro pasteles que fueron horneados en las diferentes combinaciones de temperatura y tiempo. Para ser un experimento factorial se tendrían que haber preparado en orden aleatorio los 32 pasteles uno por uno. Pero se trata de un diseño en parcelas divididas porque sólo se prepararon en orden aleatorio ocho mezclas de ingredientes (parcelas), y cada una se dividió en cuatro partes (subparcelas) que se asignaron en orden aleatorio a una condición de horneado. Note que el

Tabla 14.5 Experimento sobre el sabor de pasteles.

Factores de diseño			Factores de ruido				
			T	–1	+1	–1	+1
H	*G*	*P*	*Ti*	–1	–1	+1	+1
–1	–1	–1		1.1	1.4	1.0	2.9
+1	–1	–1		1.8	5.1	2.8	6.1
–1	+1	–1		1.7	1.6	1.9	2.1
+1	+1	–1		3.9	3.7	4.0	4.4
–1	–1	+1		1.9	3.8	2.6	4.7
+1	–1	+1		4.4	6.4	5.2	6.6
–1	+1	+1		1.6	2.1	2.3	1.9
+1	+1	+1		4.9	5.5	5.2	5.7

horno se tuvo que usar 32 veces, pero seguramente existen otras maneras más eficientes de realizar las corridas en este experimento (véase Box y Jones, 2000; Bisgaard y Kulahci, 2001). Más adelante analizamos los datos de la tabla 14.5.

Modelo y análisis estadístico de los diseños en parcelas divididas

El caso de dos factores

El caso más simple de *parcelas divididas con dos factores* es similar al ejemplo 14.3 de la tabla 14.4, donde se estudian dos pretratamientos y cuatro barnices, los primeros como parcelas y los segundos como subparcelas. En este caso los pretratamientos se aplican en un esquema completamente al azar, mientras que los barnices se aplican al azar dentro de cada pretratamiento. Este acomodo de los factores de parcela y subparcela da lugar a un análisis particular. Si en general los factores se denotan por *A* (parcela) y *B* (subparcela) con *a* y *b* niveles, respectivamente, y si consideramos *n* repeticiones de cada nivel del factor de parcela, las cuales están acomodadas en un diseño completamente al azar, el modelo estadístico que describe esta situación es:

$$y_{ijk} = \mu + \alpha_i + \delta_{ik} + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}; \quad k = 1, 2, ..., n; \; i = 1, 2, ..., a; \; j = 1, 2, ..., b \tag{14.12}$$

donde α_i es el efecto del *i*-ésimo nivel del factor de parcela y δ_{ik} es el error de parcela. Sin considerar la media general μ, estos primeros dos efectos se refieren a la parcela. Los últimos tres corresponden a la subparcela, donde β_j es el efecto del *j*-ésimo nivel del factor *B*, $(\alpha\beta)_{ij}$ es el efecto de interacción *AB* en la combinación *ij* y ε_{ijk} es el error de la subparcela. Se espera que el error de subparcela siempre sea menor que el error de parcela. El objetivo del análisis es probar las hipótesis usuales:

$$
\begin{aligned}
&H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0 \\
&H_A : \alpha_i \neq 0 \text{ para algún } i, \\
&H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0 \\
&H_A : \beta_i \neq 0 \text{ para algún } i, \\
&H_0 : (\alpha\beta)_{ij} = 0 \text{ para todo } (i, j) \\
&H_A : (\alpha\beta)_{ij} \neq 0 \text{ para algún } (i, j)
\end{aligned}
\qquad \textbf{(14.13)}
$$

Para ello es necesario obtener la tabla de ANOVA, que en los diseños en parcelas divididas es como calcular dos ANOVA; el primero para el efecto del factor de parcela y el segundo para el efecto del factor de subparcela y su interacción con el factor de parcela. De aquí que en el modelo aparezcan dos términos error, uno para cada tamaño de unidad experimental. Son precisamente los diferentes tamaños de las unidades experimentales utilizadas con cada factor lo que da lugar a las dos estructuras de error.

Para calcular las sumas de cuadrados de los efectos en el modelo se utilizan las mismas expresiones que se presentaron en los diseños con bloques (capítulo 4) y en diseños factoriales (capítulo 5), pero no las vamos a reescribir aquí. En este momento es más importante revisar la tabla de los valores esperados de cuadrados medios (tabla 14.6), donde queda en evidencia la necesidad de las dos estructuras de error mencionadas. En esta tabla también se incluyen los grados de libertad, y el factor de parcela se acomoda como en un diseño completamente al azar, con n repeticiones en cada parcela. Los factores A y B se consideran fijos y las repeticiones aleatorias. Las reglas para obtener los valores esperados de cuadrados medios se pueden consultar en Hicks (1993).

De la tabla 14.6 se concluye que la hipótesis sobre el factor A se prueba con el estadístico $F_0^A = CM_A/CM_\delta$, donde llamamos CM_δ al cuadrado medio del error de parcela completa para distinguirlo del cuadrado medio del error en la subparcela (CM_{ES}). La hipótesis acerca de B se prueba con $F_0^B = CM_B/CM_{ES}$; mientras que la hipótesis sobre la interacción AB se verifica con $F_0^{AB} = CM_{AB}/CM_{ES}$.

Análisis del ejemplo 14.3. Veamos el análisis del ejemplo 14.3, donde se quiere estudiar el efecto de dos pretratamientos (parcela) y cuatro marcas de barniz (subparcela) sobre la resistencia de la madera al agua. Vamos a proceder por pasos para apoyarnos en los ANOVA que se obtienen con cualquier software estadístico. Primero, para obtener el error de parcela sólo se toman en cuenta los promedios de cada réplica de los pretratamientos de la tabla 14.4, y se hace un ANOVA de la manera usual para comparar los pretratamientos. A las sumas de cuadrados resultantes se les multiplica por cuatro (tamaño de parcela, número de datos utilizados para calcular el promedio), para volver a la escala original de los datos. Los resultados de este primer paso se muestran en la tabla 14.7.

Enseguida, para comparar las subparcelas (barnices), primero se obtiene el ANOVA para los datos de la tabla 14.4 como si el diseño fuera factorial 2×4 con tres réplicas, véase tabla 14.8. Este ANOVA no es apropiado para el diseño de parcelas divididas, sin embargo, ayuda en la obtención del ANOVA apropiado.

Tabla 14.6 Valores esperados de cuadrados medios del diseño en parcelas divididas con las parcelas acomodadas completamente al azar.

	FV	GL	E(CM)
Parcela	α_i (A)	$a-1$	$\sigma_\varepsilon^2 + b\sigma_\delta^2 + bn\phi_A$
	δ_{ik} (error parcela)	$a(n-1)$	$\sigma_\varepsilon^2 + b\sigma_\delta^2$
Subparcela	β_j (B)	$b-1$	$\sigma_\varepsilon^2 + an\phi_B$
	$(\alpha\beta)_{ij}$ (AB)	$(a-1)(b-1)$	$\sigma_\varepsilon^2 + n\phi_{AB}$
	ε_{ijk} (error subparcela)	$a(n-1)(b-1)$	σ_ε^2
	Total	$abn-1$	

El ANOVA final del diseño en parcelas divididas se muestra en la tabla 14.9. Éste se obtiene a partir de la información de las dos tablas anteriores. De manera específica, al tomar como base la tabla 14.8, se le agrega el *error de parcela* obtenido en la tabla 14.7. El *error de subparcela* se obtiene restando al error de la tabla 14.7 el error de la tabla 14.8:

$$SC_{subparcela} = SC_E - SC_{parcela} = 927.88 - 776.72 = 151.06$$

que tiene 16 – 4 = 12 grados de libertad. Con estos datos se obtiene la tabla 14.9, donde los F_0 se consiguen dividiendo los cuadrados medios entre el error correspondiente de parcela o subparcela, según corresponda, como se desprende de la tabla 14.6. Los *valores-p* también se pueden obtener con el apoyo de un software. De la tabla 14.9 se concluye que los barnices son diferentes estadísticamente y que los pretratamientos y su interacción con los barnices no tienen un efecto considerable. De manera gráfica también es posible determinar cuál de los barnices es mejor. Vale la pena mencionar que si el experimento se hubiera analizado de manera inapropiada como si fuera un diseño factorial, de la tabla 14.8, se hubiese concluido erróneamente que los pretratamientos son importantes y que los barnices no lo son. Esto nos permite reconocer la trascendencia del diseño en parcelas divididas y su análisis apropiado.

Error de parcela
Varianza del error experimental debido a parcela, contra el que se compara el efecto de factor de parcela.

Error de subparcela
Término de error experimental debido a subparcela contra el que se compara el efecto de factor de subparcela.

Parcelas acomodadas en bloques

Si en lugar de considerar a las parcelas acomodadas en un diseño completamente al azar como en el ejemplo anterior, se consideran en un diseño en bloques completos al azar, los estadísticos de prueba cambian drásticamente. En la figura 14.3 se presenta un esquema de esta situación experimental y en el ejercicio 14.14 se muestra

Parcelas en bloques
Cuando los *k* niveles del factor de parcela se acomodan aleatoriamente dentro de bloques formados por *k* parcelas.

Tabla 14.7 ANOVA con promedio para el factor de parcela.

FV	4*SC	GL	CM	F_0	Valor-*p*
Pretratamiento	780.90	1	780.9	4.02	0.1154
Error (parcela)	776.72	4	194.18		
Total	1 557.62	5			

Tabla 14.8 Análisis del ejemplo 14.3 como factorial 2 × 4.

FV	SC	GL	CM	F_0	Valor-*p*
Pretratamiento	782.04	1	782.04	13.49	0.002
Barniz	266.00	3	88.67	1.53	0.245
PretraXBarniz	62.79	3	20.93	0.36	0.782
Error	927.88	16	57.99		
Total	2 038.72	23			

un experimento donde además de los factores parcela (variedad) y subparcela (espaciamiento de surco) está presente un factor de bloque en cuyos niveles se hacen las repeticiones.

Supongamos que en general hay dos factores de interés, *A* (parcela) y *B* (subparcela) con *a* y *b* niveles, pero ahora las *r* repeticiones de las parcelas son los niveles de un factor de bloques que denotamos por *R*. Con estas consideraciones el modelo estadístico queda como:

$$y_{ijk} = \mu + \alpha_i + r_k + (\alpha r)_{ik} + \beta_j + (\beta r)_{jk} + (\alpha\beta)_{ij} + (\alpha\beta r)_{ijk} + \varepsilon_{ijk}$$
$$i = 1, 2, ..., a;\ j = 1, 2, ..., b;\ k = 1, 2, ..., r \qquad \textbf{(14.14)}$$

donde α_i es el efecto del *i*-ésimo nivel del factor *A* (parcela), r_k es el efecto del *k*-ésimo nivel del factor *R* (bloque), $(\alpha r)_{ij}$ es el efecto de interacción *AR* en la combinación *ik*, β_j es el efecto del factor *B* en su nivel *j*, $(\beta r)_{jk}$ representa la interacción *BR* en *jk*, $(\alpha\beta)_{ij}$ es el efecto de interacción *AB* en la combinación *ij*, $(\alpha\beta r)_{ijk}$ es la interacción *ABR* en la combinación *ijk* y ε_{ijk} es el error aleatorio en la casilla *ijk*.

Observe cómo un cambio aparentemente leve en la manera de acomodar las parcelas impacta bastante los modelos (ecuaciones 14.12 y 14.14). Ahora, en lugar de tener el error de parcela se tiene el efecto del factor bloque y la interacción *AR*. Esta última será usada como error de parcela. Obviamente esto impacta los valores esperados de los cuadrados medios para estimar el efecto de cada componente del modelo, como se aprecia en la tabla 14.10.

En la tabla 14.10 queda claro que los estadísticos de prueba apropiados ahora son:

$$F_0^A = CM_A/CM_{AR};\ F_0^B = CM_B/CM_{BR} \quad \text{y} \quad F_0^{AB} = CM_{AB}/CM_{ABR}$$

Tabla 14.9 ANOVA completo del diseño en parcelas divididas, ejemplo 14.3.

	FV	SC	GL	CM	F_0	Valor-*p*
Parcela	Pretratamiento	782.04	1	782.04	4.03	0.1353
	Error (parcela)	776.72	4	194.18	15.28	
Subparcela	Barniz	266.00	3	88.67	7.04	0.0054
	PretraXBarniz	62.79	3	20.93	1.66	0.228
	Error (subparcela)	151.06	12	12.59		
	Total	2 038.61	23			

Tabla 14.10 Valores esperados de los cuadrados medios para las parcelas acomodadas en bloques.

	FV	GL	E(CM)
Parcela	R: r_k	$r-1$	$\sigma_\varepsilon^2 + ab\phi_R$
	A: α_i	$a-1$	$\sigma_\varepsilon^2 + b\sigma_{r\alpha}^2 + br\phi_A$
	AB: $(\alpha r)_{ik}$ (error)	$(a-1)(r-1)$	$\sigma_\varepsilon^2 + b\sigma_{r\alpha}^2$
Subparcela	B: β_j	$b-1$	$\sigma_\varepsilon^2 + a\sigma_{r\beta}^2 + ar\phi_B$
	BR: $(\beta r)_{jk}$	$(b-1)(r-1)$	$\sigma_\varepsilon^2 + a\sigma_{r\beta}^2$
	AB: $(\alpha\beta)_{ij}$	$(a-1)(b-1)$	$\sigma_\varepsilon^2 + a\sigma_{r\alpha\beta}^2 + r\phi_{AB}$
	ABR: $(\alpha\beta r)_{ijk}$	$(a-1)(b-1)(r-1)$	$\sigma_\varepsilon^2 + \sigma_{r\alpha\beta}^2$
	ε_{ijk} (error)	0	σ_ε^2 (no estimable)
	Total	$abr-1$	

donde todos los denominadores son cuadrados medios de interacciones que involucran las réplicas (bloques), mientras que el CM_E ni siquiera es estimable debido a que sólo hay un dato en cada combinación de (bloque, A, B). Por lo general, se combinan las sumas de cuadrados de las interacciones que involucran las repeticiones o bloques (R) para conformar un error de subparcela con $a(b-1)(r-1)$ grados de libertad ($SC_E = SC_{BR} + SC_{ABR}$).

El caso con más de dos factores

Como se mencionó antes, existen situaciones experimentales que llevan a una estructura más compleja de parcelas divididas, dependiendo principalmente de cómo se arreglen las parcelas y las subparcelas y de cuántos factores de proceso o factores de bloque se incluyan en el experimento. Una situación especial se presenta con los experimentos robustos con *arreglo interno y externo* (capítulo 9), cuando éstos en realidad son corridos como parcelas divididas, como en el ejemplo 14.4. Muchas veces, estos diseños no se corren completamente aleatorizados porque el arreglo interno (de factores de proceso) está ordenado en función de factores que son difíciles de mover, y en este caso no es conveniente hacer las corridas en orden aleatorio. De tal forma que, una vez que se fija una combinación de niveles de los factores de proceso, se obtienen en orden aleatorio los datos para todas las combinaciones de niveles de los factores de ruido. De manera que cada combinación de niveles de los factores de proceso es una *parcela* y, dentro de ésta, cada combinación de niveles de los factores de ruido es una *subparcela*.

Arreglo interno-externo en parcelas divididas
Cada combinación de los factores de control es una parcela y cada combinación de los factores de ruido forman una subparcela.

Análisis del ejemplo 14.4. Veamos el modelo y el análisis del ejemplo 14.4, donde se quiere diseñar una harina robusta, con tres factores de proceso y dos factores de ruido. Por la manera en que se corrió este experimento se trata de un diseño en parcelas dividas, en el cual los factores de proceso definen las parcelas y los factores de ruido las subparcelas. El modelo estadístico para este tipo de situaciones está dado por:

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \phi_k + (\alpha\beta)_{ij} + (\alpha\phi)_{ik} + (\beta\phi)_{jk} + (\alpha\beta\phi)_{ijk} + \delta_{ijk}$$
$$+ \varphi_l + \gamma_m + (\varphi\gamma)_{lm} + (\alpha\varphi)_{il} + (\beta\varphi)_{jl} + (\phi\varphi)_{kl} + (\alpha\gamma)_{im} + (\beta\gamma)_{jm} + (\phi\gamma)_{km} + \varepsilon_{ijklm} \quad \textbf{(14.15)}$$

con $i, j, k, l, m = 1, 2$. Los términos en el primer renglón de la ecuación son los efectos en las parcelas y los del segundo renglón son los efectos en las subparcelas. Note que los efectos de la interacción de factores de proceso con factores de ruido, además de que son de particular interés para lograr la robustez del producto, forman parte de los efectos de subparcela.

En el caso del ejemplo de la harina no hay réplicas en el experimento, por lo tanto, los términos de error de parcela δ_{ijk} y de subparcela ε_{ijklm} no son estimables en forma directa; más bien, se estimarán con los efectos no activos y con los de alto grado. Por ejemplo, ε_{ijklm} estaría conformado por las 15 interacciones de alto orden que faltan en el modelo, como ocurre en los diseños factoriales sin réplicas que se estudiaron en el capítulo 6. Interesa probar la significancia de los 16 efectos potencialmente importantes incluidos en el modelo (14.15), pero como no hay réplica, no tiene sentido hacer directamente el ANOVA de parcela grande, como se hizo en el ejemplo anterior.

En este caso, en primer lugar se recomienda estimar todos los efectos analizando los datos como si fueran un factorial completo 2^5. En el lado izquierdo de la figura 14.5 se muestran los efectos estimados; a continuación, los efectos estimados se separan en efectos de parcela y de subparcela de acuerdo con el modelo, y cada grupo se grafica por separado en papel de probabilidad normal (gráfico de Daniel). En la figura 14.5 se aprecia que los efectos de parcela que no se alinean y que parecen importantes son los efectos *H* (harina) y *P* (polvo de huevo), mientras que los efectos subparcela son *T* (temperatura), *Ti* (tiempo), *GT* (grasa × temperatura) y posiblemente *HT* (harina × temperatura).

Con base en lo anterior, los efectos no significativos en cada unidad experimental se combinan para construir los errores y así obtener ANOVA aproximados. En el caso de parcelas, como sólo *H* y *P* parecen significativos, de la tabla 14.5 se calculan los promedios obtenidos en cada una de las combinaciones de *H* y *P*, y se obtiene el ANOVA como si se tratara de un diseño 2×2 con dos réplicas. Las sumas de cuadrados obtenidas se multiplican por el número de datos con los que se obtuvo cada promedio (en este caso 4), para conservar la escala. De esta manera se obtiene el ANOVA para los efectos en la parcela completa (factores de proceso) de la tabla 14.11. Asimismo, se aprecia que, efectivamente, los efectos harina (*H*) y polvo de huevo (*P*) resultan significativos.

Sólo para facilitar los cálculos en la tabla 14.12, el ANOVA del experimento se muestra como si fuera un factorial completo 2^5, donde el término error se estima con los 24 efectos más pequeños detectados por dicho análisis. Este análisis no detecta la interacción *HT* como significativa, pero se decide dejarlo en la tabla porque el análisis gráfico lo muestra como candidato a estar activo. En cambio, la interacción *GP* se

Tabla 14.11 ANOVA para los efectos de parcela.

FV	*4*SC*	*GL*	*CM*	F_0	Valor-*p*
H (harina)	52.78	1	52.78	61.09	0.000
P (polvo de huevo)	11.64	1	11.64	13.47	0.014
Error de parcela	4.32	5	0.86		
Total	68.74	7			

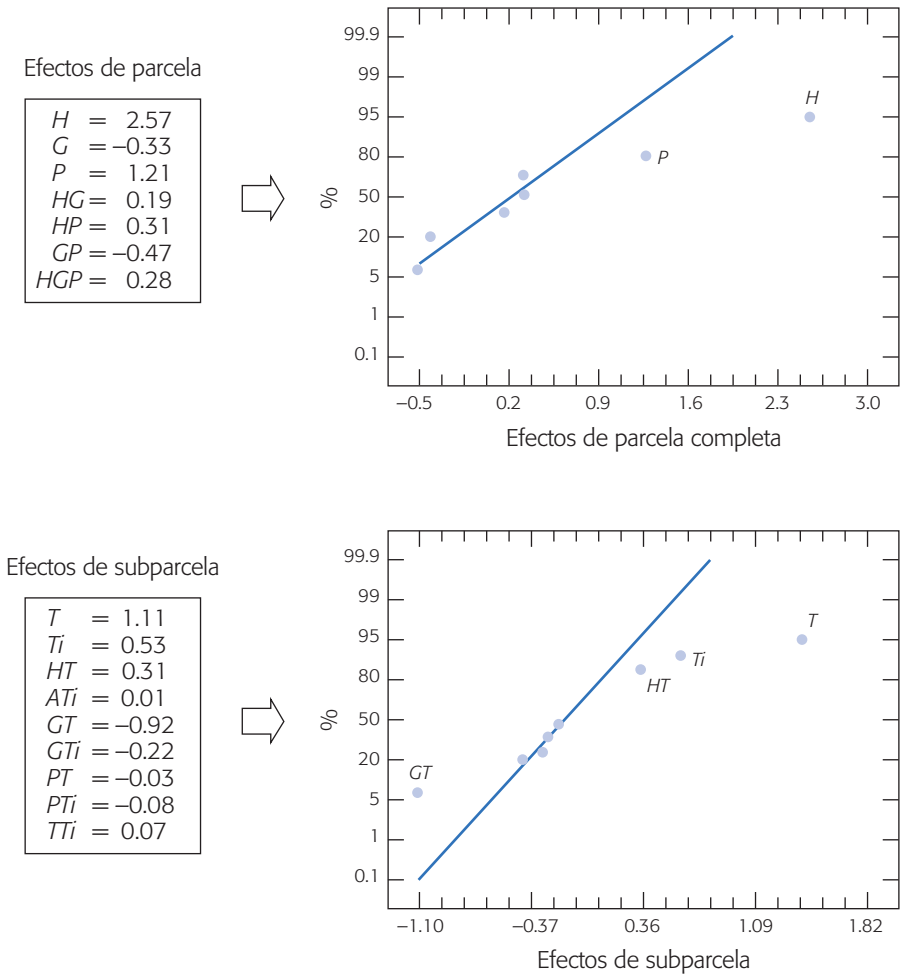


Figura 14.5 Efectos estimados para el ejemplo 14.4 y graficados en papel de probabilidad normal.

detecta como significativa, pero en el análisis gráfico no aparece como tal. El resto de los efectos de la tabla también se detectaron activos en el análisis gráfico. El ANOVA de la tabla 14.12 no es apropiado para el diseño porque el término de error no es correcto. De hecho, es típico que el error del análisis factorial subestime el error de parcela y sobreestime el error de subparcela.

Para obtener el análisis de varianza para parcelas divididas se usa como base la tabla 14.12, pero se corrige agregando el error de parcela estimado en la tabla 14.11, y también se estima el error de subparcela. Para esto, al error de la tabla 14.12 se le resta y agregan términos de la siguiente manera:

$$SC_{subparcela} = SC_E - SC_{parcela} + SC_{GP} = 6.16 - 4.32 + 1.76 = 3.6$$

que tiene 20 grados de libertad. La interacción GP se mandó al error porque en el análisis gráfico se detectó que no era significativa. En la tabla 14.13 se muestra el ANOVA completo para el diseño de parcelas divididas. Cada efecto se contrasta con el error que le corresponde. Ahora, la interacción HT, al ser contrastada con el error

Tabla 14.12 ANOVA como diseño factorial (análisis incorrecto).

FV	SC	GL	CM	F_0	Valor-*p*
H (harina)	52.78	1	52.79	205.41	0.0000
P (polvo de huevo)	11.64	1	11.64	45.29	0.0000
T (temperatura)	9.79	1	9.79	38.09	0.0000
Ti (tiempo)	2.26	1	2.26	8.79	0.0067
GT	6.75	1	6.75	26.26	0.0000
GP	1.76	1	1.76	6.85	0.0151
HT	0.75	1	0.75	2.92	0.1004
Error	6.16	24	0.257		
Total	91.90	31			

de subparcela, aumenta su significancia hasta casi 0.05, por lo que valdría la pena interpretarla. El resto de los efectos mantienen sus características. Cabe aclarar que al no existir repeticiones genuinas en el experimento se tuvieron que hacer las gráficas de efectos para construir los errores en forma apropiada mediante la suma de cuadrados de efectos pequeños. En caso de que haya repeticiones se procede en forma similar pero todo se obtiene directamente de los ANOVA.

Como siempre, la interpretación de los efectos activos se hace de manera gráfica como en los diseños factoriales. En la figura 14.6 sólo representamos las interacciones que resultaron significativas en las subparcelas, y que involucran un factor de proceso y un factor de ruido. Con estas interacciones es posible hacer un intento para lograr la robustez de la harina, eligiendo el nivel del factor de control que minimiza el efecto del factor de ruido. De la interacción *GT* se aprecia que en el nivel alto de *G* (grasa) hay poco efecto de la variación de la temperatura sobre el sabor de los pasteles. En cambio, la interacción *HT* es muy leve y no hay mucha diferencia en el efecto de *T* en los diferentes niveles de *H*. Por ello, se recomienda un nivel alto de *H* para atender el sabor. Para fijar un nuevo nivel del otro factor de proceso significativo (*P*) se recomienda ver su gráfica de efectos principales en donde se observa que lo más conveniente es el nivel alto.

Tabla 14.13 Análisis completo del diseño en parcelas divididas, ejemplo 14.4.

	FV	SC	GL	CM	F_0	Valor-*p*
Parcela	*H* (harina)	52.78	1	52.78	61.37	0.000
	P (polvo de huevo)	11.64	1	11.64	13.53	0.014
	Error de parcela	4.32	5	0.86	4.78	
Subparcela	*T* (temperatura)	9.79	1	9.79	54.39	0.000
	Ti (tiempo)	2.26	1	2.26	12.55	0.002
	GT	6.75	1	6.75	37.5	0.000
	HT	0.75	1	0.75	4.17	0.054
	Error de subparcela	3.6	20	0.18		
	Total	91.90	31			

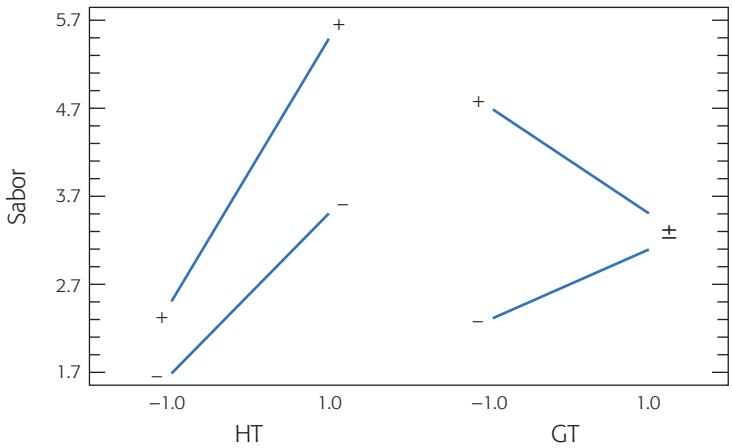


Figura 14.6 Gráficos de interacción.

Cómo hacer los cálculos usando software

Todos los cálculos se pueden generar a partir de las opciones de análisis de experimentos factoriales que se estudiaron en los capítulos previos; sin embargo, es más fácil cuando se usa un software especializado. El diseño completamente anidado se puede analizar en *Minitab* usando la secuencia: *Stat* → *ANOVA* → *Fully Nested ANOVA*. Pero si también se tienen factores cruzados (no anidados) se usa la opción: *Stat* → *ANOVA* → *Balanced ANOVA*. Es preciso especificar en la notación del propio software el modelo que se quiere utilizar, que distingue entre términos cruzados y anidados. Por ejemplo, el ANOVA de la tabla 14.3 se obtiene declarando el modelo (ecuación 14.10) como: ENSAM ARREG → ENSAM*ARREG → OPERA(ARREG) → ENSAM* OPERA(ARREG). También se declara cuáles factores son aleatorios, ya que de lo contrario el software supondrá que los factores son fijos. En las versiones recientes de *Statgraphics* los diseños anidados se incluyen en *Compare* → *Analysis of variance* → *General linear model*; después de introducir los factores aparecerá una pantalla donde se define la estructura de anidamiento y si alguno de ellos es aleatorio.

El análisis de parcelas divididas se realiza usando las opciones usuales para diseños factoriales con dos o más niveles, ya sea en *Statgraphics* o *Minitab*. De manera alternativa se usa la opción más general de *ANOVA* → *Compare* → *Analysis of variance* → *Multifactor ANOVA*, en *Statgraphics*; o bien, *Stat* → *ANOVA* → *Balanced ANOVA*, en *Minitab*. A diferencia de *Minitab*, en *Statgraphics* de entrada sólo se declaran los efectos principales en *Multifactor ANOVA*; después, la interacción se pide en opciones de análisis (*Analysis options*) con el botón derecho del puntero ("ratón").

Preguntas y ejercicios

1. Explique en qué consiste un diseño anidado y proporcione un ejemplo.
2. ¿Por qué en un diseño anidado con dos factores no se puede estudiar la interacción de éstos?

3. Bosqueje la tabla de ANOVA para un diseño con tres factores *A*, *B* y *C*, con *B* anidado en *A* y *C* anidado en *B*. Escriba el modelo estadístico asociado.
4. ¿Qué es un diseño en parcelas divididas? Escriba un ejemplo.
5. Describa un par de situaciones prácticas que llevan a utilizar un diseño en parcelas divididas.
6. Explique las diferencias entre un diseño en parcelas divididas y un diseño anidado, ambos con dos factores.

Ejercicios de diseños anidados

7. Con respecto a los datos del ejemplo 14.1 dados en la figura 14.1:
   a) Analice los datos como si fuera un estudio R&R cruzado o factorial. Compare la tabla de ANOVA resultante con la tabla 14.2, la cual representa el análisis anidado correcto. Verifique los cálculos de la $SC_{B(A)}$ usando la relación dada en la ecuación 14.8.
   b) Si el estudio se analiza como un diseño factorial, ¿empeora o mejora de la calidad del sistema de medición? Para cada análisis obtenga los componentes de varianza y calcule qué porcentaje de la variabilidad total se atribuye al sistema de medición.
8. Con respecto a los datos del ejemplo 14.2 dados en la figura 14.2.
   a) Analice los datos considerando el factor operador fijo, después compare los resultados con el análisis dado en la tabla 14.3. Por último, explique las diferencias observadas en los valores-*p*.
   b) Haga otra vez el análisis pero ahora como si fuera un experimento factorial incluyendo hasta la interacción triple. A partir de este análisis obtenga las sumas de cuadrados anidadas de la tabla 14.3, usando las ecuaciones (14.11).
9. Suponga que le muestran la tabla de ANOVA que aparece a continuación, en la cual un diseño anidado se analizó como un diseño factorial (cruzado). Reconstruya la tabla de ANOVA correcta asumiendo que *B* está anidado en *A* (*A* es fijo, *B* es aleatorio).

FV	*GL*	*SC*	*CM*	F_0	Valor-*p*
A	2	15	7.5	2.50	0.10
B	3	24	8.0	2.67	0.07
AB	6	42	7.0	2.33	0.06
Error	24	72	3.0		
Total	35	153			

10. Analice otra vez los datos del experimento anidado que se presentó en la figura 14.1, pero ahora considere ambos factores fijos. Obtenga la tabla de ANOVA del diseño anidado a partir de la tabla de ANOVA del diseño factorial. Compare los resultados con los reportados en la tabla 14.2.
11. Beckman *et al*. (1987) describen un experimento cuyo objetivo es determinar si el aerosol estándar utilizado para probar los filtros de respiración puede reemplazarse por un aerosol alternativo (aerosol 2). También interesaba investigar la variabilidad de los filtros de dos fabricantes. Los datos obtenidos, donde la respuesta es el porcentaje de penetración del aerosol, se muestran a continuación:

Fabricante	Filtro	Aerosol 1			Aerosol 2		
1	1	0.750	0.770	0.840	1.120	1.100	1.120
	2	0.082	0.076	0.077	0.150	0.120	0.120
2	1	0.600	0.680	0.870	0.910	0.830	0.950
	2	1.000	1.800	2.700	2.170	1.520	1.580

a) Escriba el modelo que plantea este experimento considerando el anidamiento que ocurre.
b) Haga el análisis y saque conclusiones.
c) Considere que interesa minimizar la respuesta, ¿recomendaría usted el aerosol 2 como alternativa para probar los filtros?

12. Se realiza un experimento para estudiar las concentraciones de calcio en plantas de nabo. Para ello, se eligen cuatro plantas al azar, de cada planta se seleccionan al azar tres hojas y de cada hoja se sacan dos muestras a las que se les mide la concentración de calcio, con los siguientes resultados:

Muestra	Planta 1			Planta 2			Planta 3			Planta 4		
	Hoja 1	Hoja 2	Hoja 3	Hoja 1	Hoja 2	Hoja 3	Hoja 1	Hoja 2	Hoja 3	Hoja 1	Hoja 2	Hoja 3
*M*1	3.28	3.52	2.88	2.46	1.87	2.19	2.77	3.74	2.55	3.78	4.07	3.31
*M*2	3.09	3.48	2.80	2.44	1.92	2.19	2.66	3.44	2.55	3.87	4.12	3.31

a) Los tres factores (planta, hoja y muestra) están completamente anidados. Explique por qué.
b) Analice los datos como un diseño completamente anidado con los tres factores aleatorios y saque conclusiones. ¿Hay diferencias entre las plantas? ¿Hay diferencias entre las hojas?

13. Un genetista colecta tres semillas de dos árboles en cada uno de los tres bosques seleccionados. Las semillas se siembran en un vivero y se mide su crecimiento. Los datos en pulgadas son los siguientes:

Árbol	Bosque		
	A	*B*	*C*
1	15.8	18.5	12.3
	15.6	18.0	13.0
	16.0	18.4	12.7
2	13.9	17.9	14.0
	14.2	18.1	13.1
	13.5	17.4	13.5

a) Plantee un modelo estadístico que considere el anidamiento entre los dos factores.
b) Analice los datos y saque conclusiones. Explique y pruebe las hipótesis de interés.
c) Analice los datos como diseño factorial y compare los resultados con el análisis del inciso anterior.

Ejercicios de parcelas divididas

14. En Steel y Torrie (1980) se publicó el siguiente experimento, que estudia el efecto de cuatro diferentes espaciamientos entre surcos sobre el rendimiento de dos variedades de soya. Los rendimientos en fanegas por acre se muestran en la siguiente tabla. Una fanega es igual 55.5 litros, y un acre es igual a 4 047 m^2.

Variedad	Espacio en pulgadas	Bloque					
		1	2	3	4	5	6
1	18	33.6	37.1	34.1	34.6	35.4	36.1
	24	31.1	34.5	30.5	32.7	30.7	30.3
	30	33.0	29.5	29.2	30.7	30.7	27.9
	36	28.4	29.9	31.6	32.3	28.1	26.9
	42	31.4	28.3	28.9	28.6	18.5	33.4
2	18	28.0	25.5	28.3	29.4	27.3	28.3
	24	23.7	26.2	27.0	25.8	26.8	23.8
	30	23.5	26.8	24.9	23.3	21.4	22.0
	36	25.0	25.3	25.6	26.4	24.6	24.5
	42	25.7	23.2	23.4	25.6	24.5	22.9

a) Describa mediante un dibujo la manera de aleatorizar este experimento. Considere a la variedad como el factor de parcela, que está acomodado en un diseño en bloques, y al espaciamiento entre surcos como el factor de subparcela.
b) Escriba el modelo estadístico y las hipótesis de interés.
c) Obtenga la tabla de análisis de varianza e interprete los resultados.

15. Potcner y Kowalski (2004) describen un experimento con un factor difícil de mover (*D*) y tres factores fáciles de mover (*A*, *B* y *C*), cada uno de éstos con dos niveles. Los niveles del factor *D* son las parcelas que se repiten dos veces, y dentro de cada parcela se tienen ocho subparcelas que son las combinaciones de niveles de los factores *A*, *B* y *C*. El experimento y los datos obtenidos se muestran en la siguiente tabla:

D	*A*	*B*	*C*	*Y*	*D*	*A*	*B*	*C*	*Y*
1	–1	1	1	108.4	1	–1	1	1	100.8
1	1	–1	1	131.6	1	1	1	–1	114.4
1	–1	–1	–1	124.0	1	1	–1	1	132.8
1	1	–1	–1	134.9	1	1	–1	–1	131.4
1	–1	1	–1	103.7	1	–1	–1	–1	118.4
1	1	1	–1	112.9	1	–1	1	–1	104.4
1	1	1	1	113.4	1	1	1	1	111.7
1	–1	–1	1	122.3	1	–1	–1	1	121.1
–1	–1	–1	–1	119.3	–1	1	1	–1	116.7
–1	1	1	–1	120.9	–1	–1	1	–1	112.8
–1	1	1	1	123.0	–1	–1	1	1	112.2
–1	1	–1	1	127.9	–1	1	–1	1	127.7
–1	–1	1	1	117.3	–1	–1	–1	–1	118.4
–1	–1	–1	1	120.9	–1	1	1	1	120.9
–1	1	–1	–1	129.9	–1	1	–1	–1	127.0
–1	–1	1	–1	115.4	–1	–1	–1	1	119.4

a) Describa el procedimiento de aleatorización que se debe seguir con este diseño.
b) Escriba el modelo estadístico y bosqueje la tabla de ANOVA.
c) Obtenga la tabla de ANOVA para el efecto en la parcela analizando el promedio por parcela. Ajuste las sumas de cuadrados a la escala de los efectos originales.
d) Haga el análisis incorrecto como si fuera un diseño factorial e interprételo.
e) Combinando de forma manual los análisis de los incisos *c*) y *d*), obtenga el análisis como un diseño en parcelas divididas.
f) Interprete y compare los resultados con el análisis incorrecto del inciso *d*). ¿Qué efectos se pueden malinterpretar cuando el experimentador sólo realiza ese análisis?
g) Interprete los efectos activos suponiendo que interesa maximizar la respuesta.

16. Considerando el experimento descrito en el ejemplo 14.4. Conteste lo siguiente.
a) Analice los datos pero ahora considerando a los factores de ruido como las parcelas y a los factores de proceso como las subparcelas.
b) ¿Cómo se habría corrido el experimento en el caso planteado en el inciso *a*)?
c) Compare sus resultados con el análisis de la tabla 14.13.

17. En un laboratorio se usan tres hornos para correr un experimento cuyo fin es estudiar el efecto de la temperatura y la orientación sobre la resistencia de aleaciones de acero. Cada horno tiene diferente temperatura; dentro de cada uno se colocan orientadas al azar (orientación 1) dos muestras de cada una de las tres aleaciones, y otras dos muestras de las mismas aleaciones se colocan alineadas (orientación 2). Es decir, dentro de cada horno se colocan en lugares elegidos al azar 12 muestras, seis orientadas al azar y seis alineadas. Los datos obtenidos se muestran en la siguiente tabla.

Orientación	Temperatura								
	675			700			725		
	Alea 1	Alea 2	Alea 3	Alea 1	Alea 2	Alea 3	Alea 1	Alea 2	Alea 3
1	12 19	15 28	20 26	27 40	35 39	48 55	47 55	55 63	62 58
2	15 23	25 31	25 33	48 55	48 62	55 64	48 60	63 68	68 62

a) ¿Cuál es el factor de parcela y cuáles son los factores de subparcela? ¿En dónde radica la restricción a la aleatorización en este experimento?
b) Escriba un modelo estadístico para este experimento.
c) Analice los datos y saque conclusiones. ¿Por qué no es posible probar el efecto individual de la temperatura?
d) Determine la mejor combinación de temperatura, orientación y aleación pensando que interesa maximizar la resistencia. Utilice gráficas.

18. Considere un experimento de horneado donde se quiere estudiar el efecto de tres temperaturas y tres tiempos de horneado sobre el sabor del producto; el arreglo propuesto es el siguiente:

Temperatura	Tiempo		
	5 min.	7 min.	9 min.
175 °C			
190 °C			
205 °C			

a) Si el experimento se corriera como un diseño factorial, explique cómo se deben obtener los datos para llenar la tabla.

b) Si se corriera como diseño en parcelas divididas, explique la manera de obtener los datos. ¿Cuál de los dos factores tomaría usted como factor de parcela? ¿Qué se ganaría con respecto al diseño factorial?

c) En los dos diseños mencionados bosqueje las tablas de ANOVA correspondientes.

Capítulo 15

Diseño de experimentos con mezclas

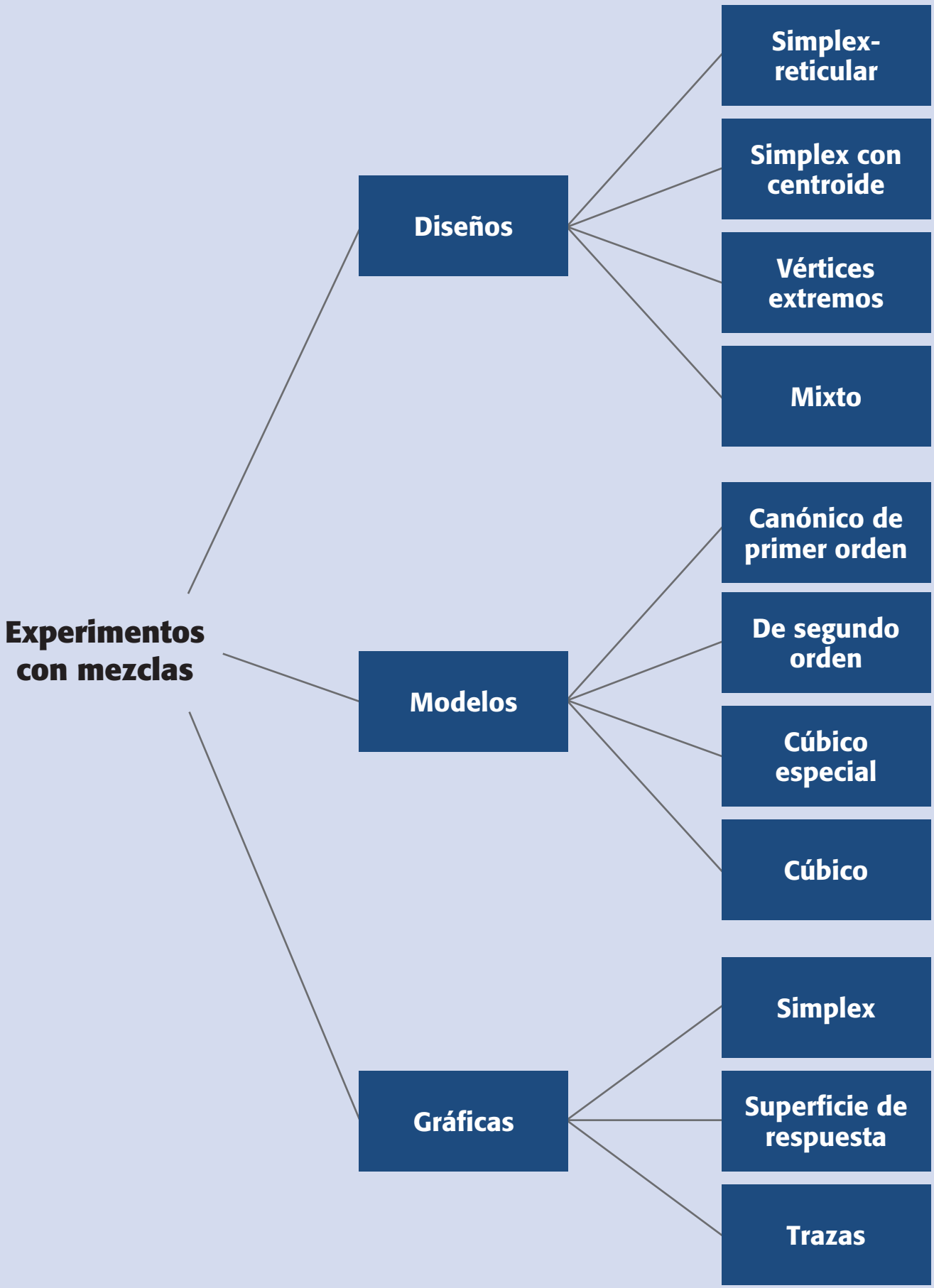
Sumario

- El problema del diseño de experimentos con mezclas
- Algunos diseños de mezclas y sus modelos estadísticos
- Ajuste del modelo y caracterización de la superficie de respuesta
- Restricciones en los componentes de una mezcla
- Uso de software estadístico

Objetivos de aprendizaje

- Comprender la naturaleza de la problemática de los experimentos con mezclas.
- Identificar los principales diseños de experimentos para mezclas junto con su representación geométrica y los modelos estadísticos adecuados para analizar los resultados experimentales.
- Ser capaz de diseñar y analizar un experimento con mezclas.

Mapa conceptual



Conceptos clave

- Coeficiente lineal β_i
- Diseño de vértices extremos
- Diseño simplex con centroide
- Diseño simplex-reticular $\{q, m\}$
- Experimento mixto
- Experimentos con mezclas
- Gráfico de trazas
- Problemas de mezclas
- Seudocomponentes
- Simplex

Experimentos con mezclas
Diseños en los que factores son los componentes o ingredientes de una mezcla. Las variables de respuesta dependen de las proporciones con las que participan los ingredientes en la mezcla y no de la cantidad de mezcla.

En los capítulos anteriores se estudiaron algunos diseños en donde los niveles de cada factor son independientes de los niveles de los otros factores. Sin embargo, en los *experimentos con mezclas*, los factores son los componentes o ingredientes de una mezcla y, como se verá más adelante, los niveles de dichos ingredientes no son independientes. Este hecho hace que los diseños que se explicaron en capítulos anteriores no se apliquen a experimentos con mezclas. En este capítulo abordaremos los aspectos fundamentales del diseño y análisis de experimentos con mezclas.

El problema del diseño de experimentos con mezclas

Existen muchos problemas reales y de investigación que involucran productos que resultan al mezclar diferentes componentes. Por ejemplo: bebidas, medicamentos, detergentes, pinturas, resinas, gasolinas, cementos, etc., están formados por una mezcla de distintos ingredientes o componentes. En general, se supone que las características de calidad de la mezcla dependen de las proporciones con las que participan los ingredientes y no de la cantidad absoluta de ellos.

Entre los objetivos de un experimento con mezclas se encuentran:

- Determinar cuáles de los ingredientes de la mezcla o interacciones entre ellos tienen mayor influencia sobre una o varias respuestas de interés.
- Modelar las respuestas de interés en función de las proporciones de los componentes de la mezcla.
- Usar dichos modelos para determinar en qué porcentaje debe participar cada uno de los ingredientes para lograr que la fórmula tenga las propiedades deseadas.

Ejemplo 15.1

En Lobato Calleros *et al*. (1997) se presentan los resultados de una investigación cuyo propósito fue generar quesos con grasas vegetales. Cada tratamiento consistió en elaborar tres kilos de análogo de queso, donde la porción grasa contribuye con 765.2 g. La fracción no grasa se dejó fija; en tanto, la fracción grasa, que es objeto de investigación, estuvo integrada por distintas proporciones de grasa butírica (x_1), grasa de soya (x_2) y aceite de soya (x_3). Se midieron diversas variables de respuesta, entre ellas dureza (y_1, kg) y elasticidad (y_2, cm). Los resultados promedio de tres réplicas se muestran en la tabla 15.1. Como se aprecia en esa tabla, los tratamientos (mezclas probadas) están definidos por la proporción en que interviene cada tipo de grasa para formar la fracción grasa del queso. Por ejemplo, la mezcla (x_1, x_2, x_3) = (1, 0, 0) significa que 100% de la porción grasa se generó con grasa butírica. Mientras que la mezcla (x_1, x_2, x_3) = (0.333, 0.333, 0.333) significa que los 765.2 g de la porción grasa del análogo de queso se forman en partes iguales por las tres grasas. El objetivo es saber qué componentes (grasa) influyen más en las características del queso. También se quiere conocer cuál es la combinación óptima si se desea lograr cierta dureza y elasticidad.

Tabla 15.1 Mezclas y resultados para el ejemplo de análogos de queso.

x_1	x_2	x_3	y_1	y_2
1	0	0	0.32	0.97
0	1	0	0.7	0.83
0	0	1	0.2	0.87
0.5	0.5	0	0.33	0.7
0.5	0	0.5	0.23	0.82
0	0.5	0.5	0.27	0.74
0.333	0.333	0.333	0.31	0.78
0.667	0.167	0.167	0.32	0.76
0.167	0.667	0.167	0.49	0.73
0.167	0.167	0.667	0.23	0.82

Al planear un experimento con mezclas se decide, en función de lo que se conoce del problema y del objetivo, las mezclas que se probarán para que al analizar los resultados se logren responder las preguntas planteadas. Es decir, no se trata de producir las formulaciones que se le ocurran al experimentador y llevar a cabo una investigación a prueba y error, sino aplicar una metodología de planeación y análisis que asegure obtener conocimiento y soluciones. Por ejemplo, las mezclas de la tabla 15.1 se decidieron con base en un diseño *simplex-reticular* (*simplex-lattice*) aumentado que, como se verá más adelante, es un diseño adecuado para muchas situaciones prácticas.

Problema de mezclas
Se tienen q ingredientes y cada tratamiento consiste de una combinación particular de ellos ($x_1, x_2, \ldots, x_q$). Se cumple que $0 \leq x_i \leq 1$ y $x_1 + x_2 + \ldots + x_q = 1$.

En general, en un problema de experimentos con mezclas se tendrán q componentes o ingredientes y cada tratamiento en el experimento consiste en una combinación particular o mezcla de dichos ingredientes. Si se denotan por $x_1, x_2, \ldots, x_q$, las proporciones en las que participan los componentes de la mezcla deben satisfacer dos restricciones:

$$0 \leq x_i \leq 1, \text{ para cada componente } i$$

$$\sum_{i=1}^{q} x_i = x_1 + x_2 + \ldots + x_q = 1$$

La primera indica que las proporciones tienen que ser cantidades entre cero y uno, y la segunda condiciona a que las q proporciones sumen siempre la unidad, lo cual causa que los niveles de los componentes x_i no sean independientes entre sí. De aquí que los diseños de experimentos con mezclas sean diferentes de los diseños de experimentos factoriales y de los diseños de superficie de respuesta usuales. Por ejemplo, en un diseño factorial donde uno de los factores es temperatura, los niveles de este factor se pueden mover de manera independiente de los demás factores. Mientras que en un diseño de mezclas, al aumentar la proporción de un ingrediente necesariamente se reduce la participación de los componentes restantes en esa misma proporción.

Simplex
Son figuras que permiten dibujar q componentes usando $q-1$ dimensiones. Se utilizan para representar geométricamente las regiones experimentales en experimentos con mezclas.

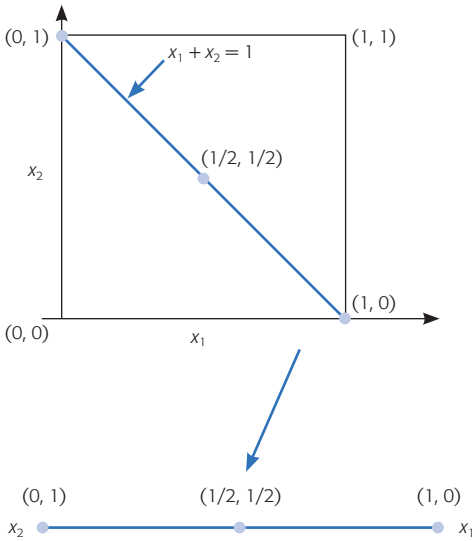
Para entender mejor el problema de los experimentos con mezclas veamos cómo se representan sus regiones experimentales mediante *simplejos* (simplex), es decir, figuras geométricas que permiten dibujar q componentes usando $q - 1$ dimensiones. En la figura 15.1 se muestra la región experimental para $q = 2$ y $q = 3$ componentes. Para dos componentes, la región experimental viable está dada por la línea recta $x_1 + x_2 = 1$, que se representa en la figura 15.1*a*) con el segmento de línea que une los puntos (1, 0) y (0, 1) en el plano cartesiano. Para $q = 3$ componentes, la región está formada por el plano que satisface la relación $x_1 + x_2 + x_3 = 1$, de donde se desprende que la región experimental es un triángulo equilátero (figura 15.1*b*)). Los tres vértices corresponden a las mezclas puras (formadas por un solo ingrediente), los tres lados o aristas representan las mezclas binarias que sólo tienen dos de los tres componentes. Los puntos interiores del triángulo representan mezclas ternarias en las cuales los tres ingredientes son diferentes de cero.

Para representar de manera adecuada cualquier mezcla ternaria es útil trazar líneas perpendiculares, a cada uno de los lados, que unan cada vértice con el lado opuesto. En la parte baja de figura 15.1*b*) cada una de estas líneas se dividió con incrementos de 0.1. En la figura 15.2 se muestra la representación geométrica para un diseño de mezclas con cuatro componentes, que corresponde a un *tetraedro*.

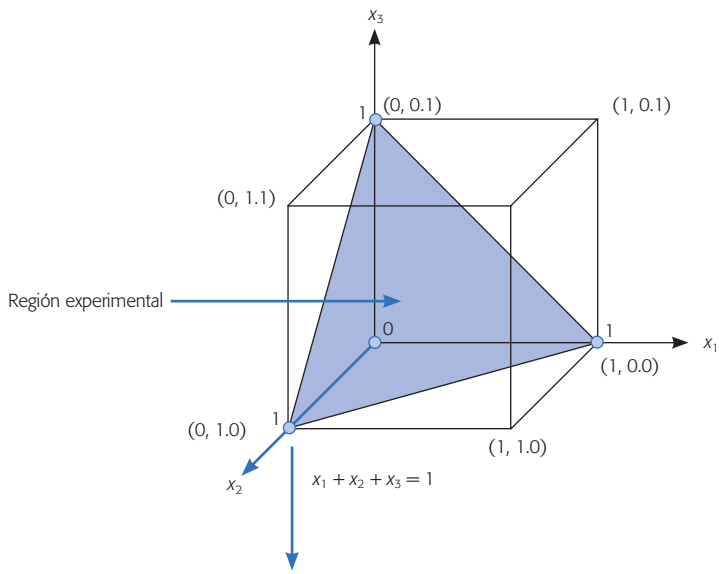
En resumen, para dos componentes la región experimental es un segmento de recta, para tres componentes es un triángulo, para cuatro componentes es un tetraedro y para más de cuatro componentes es un *hipertetraedro*. Note que en diseño de mezclas se puede dibujar con un tetraedro la región experimental para cuatro componentes, mientras que en los diseños factoriales no es posible dibujar la región experimental para cuatro factores en una figura compacta. Además, en los diseños de mezclas las regiones mencionadas son al mismo tiempo las regiones de operabilidad, ya que cualquier mezcla posible es un punto del simplex.

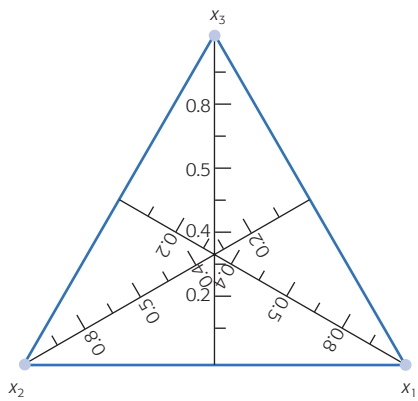
Los diseños de experimentos que se utilizan en mezclas distribuyen adecuadamente los tratamientos o mezclas en la región experimental. Por ejemplo, la representación de los tratamientos (mezclas) en el ejemplo de análogos de queso de la tabla 15.1 se muestra en la figura 15.3*b*). Con frecuencia la investigación se restringe a una parte de la región experimental, y esto ocurre cuando por restricciones físicas o conocimientos técnicos se quiere que uno o más de los componentes tome valores en un rango más estrecho. Por ejemplo, entre 20 y 40%. Esto lo veremos más adelante con detalle.

En general, el objetivo del diseño de experimentos con mezclas es cuantificar la influencia que tienen los diferentes componentes sobre la respuesta, tanto en forma individual como en su acción conjunta con otros componentes. Se trata de modelar esta respuesta para predecirla en cualquier formulación posible, y utilizar los modelos con el propósito de encontrar la composición de la mezcla que proporcione mejores resultados (optimizar). Para cumplir con estas tareas es preciso tomar en cuenta las particularidades de los experimentos con mezclas, ya que los diseños factoriales tradicionales no son aplicables, los polinomios estándar (modelos) no son adecuados (pues algunos de sus parámetros carecerán de sentido) y en cuanto al análisis, las pruebas estadísticas tradicionales de los modelos pueden generar confusiones.



a) Región experimental con $q = 2$.





b) Región experimental con $q = 3$.

Figura 15.1 Representación geométrica de un diseño de experimentos con mezclas.

- Los vértices corresponden a "mezclas" puras
- Las aristas a mezclas binarias
- Las caras corresponden a mezclas ternarias
- Los puntos interiores a mezclas cuaternarias

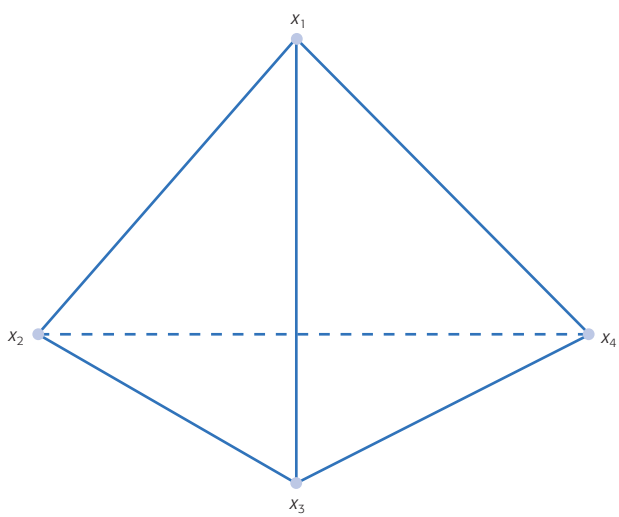


Figura 15.2 Región experimental de un diseño de mezclas con $q = 4$ componentes.

Algunos diseños de mezclas y sus modelos estadísticos

Los dos tipos básicos de diseños para estudiar el efecto de los componentes de la mezcla sobre la respuesta son el *simplex-reticular* (*simplex-lattice*) y el *simplex con centroide* (simplex-*centroide*). El diseño simplex reticular $\{q, m\}$ considera q componentes y permite ajustar un modelo estadístico de orden m. Los puntos del diseño consisten en todas las posibles combinaciones de componentes o mezclas que se forman al considerar que las proporciones pueden tomar los $m + 1$ valores entre cero y uno dados por:

$$x_i = 0,\ 1/m,\ \ 2/m,\ \ldots,\ m/m$$

Por ejemplo, un diseño simplex reticular $\{3, 2\}$ implica que $q = 3$ y $m = 2$; por ello, los valores que pueden tomar los componentes son $x_i = 0, ½, 1$. Con lo que las mezclas de tres componentes que se pueden formar con estos valores son:

$$(x_1, x_2, x_3) = (1, 0, 0);\ (0, 1, 0);\ (0, 0, 1);\ (1/2, 1/2, 0);\ (1/2, 0, 1/2) \text{ y } (0, 1/2, 1/2)$$

que corresponden a las tres mezclas puras y a tres binarias. La representación gráfica de este diseño se muestra en la figura 15.3*a*).

El diseño simplex reticular $\{3, 3\}$, implica que $q = 3$ y $m = 3$, y los valores que pueden tomar los componentes son $x_i = 0, 1/3, 2/3, 1$. Las mezclas de tres componentes que se pueden formar con estos valores son:

$$(x_1, x_2, x_3) = (1, 0, 0);\ (0, 1, 0);\ (0, 0, 1);\ (2/3, 1/3, 0);\ (2/3, 0, 1/3);\ (0, 2/3, 1/3);\ (1/3, 2/3, 0);\ (1/3, 0, 2/3);\ (0, 1/3, 2/3) \text{ y } (1/3, 1/3, 1/3)$$

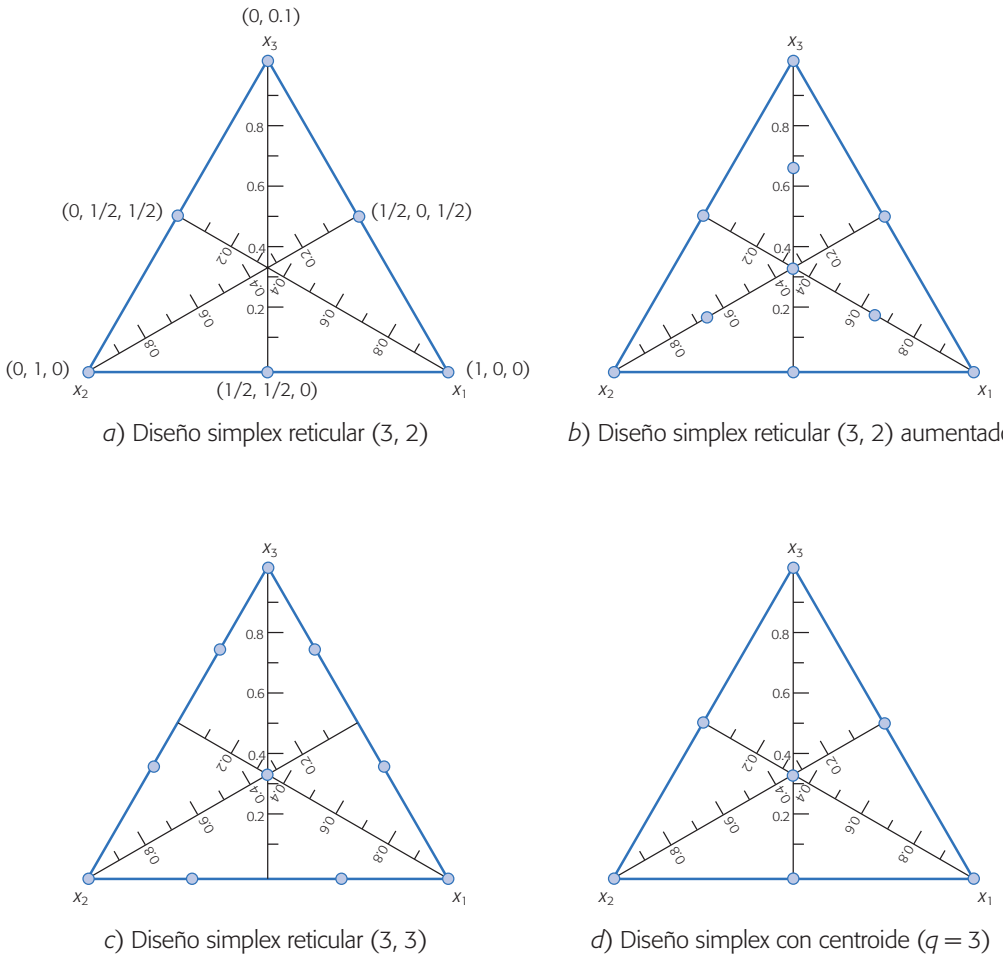
La representación gráfica de este diseño se muestra en la figura 15.3*c*). El diseño simplex reticular $\{4, 2\}$ implica que $q = 4$ y $m = 2$, y los valores que pueden tomar los componentes son $x_i = 0,\ 1/2,\ 1$. Las mezclas serán del tipo: $(x_1, x_2, x_3, x_4) = (1, 0, 0, 0)$ y $(1/2, 1/2, 0, 0)$, que corresponden a los vértices y centros de las aristas de un tetraedro.

Aumento del diseño

Diseño simplex reticular {*q, m*}
Considera q componentes y permite ajustar un modelo de orden m. Los tratamientos se obtienen con todas las mezclas que pueden formarse dado que los componentes toman los valores $x_i = 0, 1/m, 2/m, \ldots, m/m$.

El *diseño simplex reticular* incluye básicamente puntos en la frontera, pero si el experimentador desea hacer predicciones en el interior es recomendable agregar corridas que estén en el interior. En particular se recomienda agregar el centroide global y las mezclas localizadas entre el centroide y los vértices. Por ejemplo, en la figura 15.3*b*) se muestra un diseño simplex reticular $\{3, 2\}$ aumentado. Las coordenadas o mezclas de este diseño son las que se muestran en la tabla 15.1 para el ejemplo de análogos de queso.

El otro diseño básico es el simplex con centroide que se aplica con pocos componentes y consiste en 2^{q-1} puntos definidos de la siguiente forma: las q mezclas puras, todos los puntos medios de las aristas definidas por cada dos vértices del sim-



a) Diseño simplex reticular (3, 2)

b) Diseño simplex reticular (3, 2) aumentado

c) Diseño simplex reticular (3, 3)

d) Diseño simplex con centroide ($q = 3$)

Figura 15.3 Ejemplos de diseños simplex reticular y simplex con centroide con tres componentes.

plex (½, ½, 0, 0),..., (0, 0,.., ½, ½); los centroides de las caras definidas por cada tres vértices del simplex: (1/3, 1/3, 1/3, 0,..., 0), ..., (0,..., 0, 1/3, 1/3, 1/3); y así hasta obtener el centroide global ($1/q$,..., $1/q$). Por ejemplo, el simplex con centroide para $q = 3$ incluye las mezclas:

$$(x_1, x_2, x_3) = (1, 0, 0); (0, 1, 0); (0, 0, 1); (1/2, 1/2, 0); (1/2, 0, 1/2); (0, 1/2, 1/2) \text{ y } (1/3, 1/3, 1/3)$$

En la figura 15.3*d*) se muestra este diseño.

Diseño simplex con centroide

Consiste de los 2^{q-1} tratamientos siguientes: las q mezclas puras, los puntos medios de la forma (½, ½, 0, 0),..., (0, 0,..., ½, ½); los centroides de las caras de cada 3 componentes: (1/3, 1/3, 1/3, 0,..., 0), ..., (0,... 0, 1/3, 1/3, 1/3); y así hasta obtener el centroide global ($1/q$,..., $1/q$).

Modelo de primer orden

Una vez obtenidos los resultados experimentales de un diseño de mezcla es necesario ajustar un modelo estadístico para investigar el efecto de los componentes sobre la respuesta. Consideremos ahora el tipo de modelo que puede ser ajustado. Una primera aproximación sería ajustar un modelo de primer orden:

$$E(y) = \beta_0 + \sum_{i=1}^{q} \beta_i x_i \tag{15.1}$$

donde $E(y)$ es el valor esperado de la variable de respuesta y. Sin embargo, debido a la restricción $x_1 + x_2 + \ldots + x_q = 1$ es posible demostrar que los parámetros $\beta_0, \beta_1, \ldots, \beta_q$ no son únicos. Se podría hacer la sustitución:

$$x_q = 1 - \sum_{i=1}^{q-1} x_i$$

en la expresión (15.1), y de esa manera eliminar la dependencia entre los componentes. Con ello es posible estimar los parámetros $\beta_0, \beta_1, \ldots, \beta_{q-1}$. El inconveniente de esto es que no se puede estimar el efecto de q-ésimo componente porque el término $\beta_q x_q$ no estaría incluido en la ecuación. La alternativa más usual es multiplicar algunos de los términos de (15.1) por la identidad $x_1 + x_2 + \ldots + x_q = 1$. Por ejemplo, si multiplicamos a β_0 por esta identidad y agrupamos términos se obtiene:

$$E(y) = \beta_0 (x_1 + x_1 + \ldots + x_1) + \sum_{i=1}^{q} \beta_i x_i = \sum_{i=1}^{q} \beta_i^* x_i \tag{15.2}$$

donde $\beta_i^* = \beta_0 + \beta_i$ para todo $i = 1, 2, \ldots, q$. El modelo con los coeficientes β_i^* se conoce como *forma canónica del modelo de primer orden para mezclas*. En la práctica, para simplificar la notación se elimina el asterisco de cada parámetro y el modelo canónico de primer orden se denota con:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i \tag{15.3}$$

Modelo cuadrático

Cuando se ajusta un modelo cuadrático también es necesario incorporar la restricción $x_1 + x_2 + \ldots + x_q = 1$, ya que esto le dará una característica especial al modelo. Para ilustrar la idea supongamos que se tienen tres componentes x_1, x_2, x_3, por lo que el polinomio de segundo grado está dado por:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 \tag{15.4}$$

Los tres términos cuadráticos se pueden expresar como,

$$x_i^2 = x_i \left(1 - \sum_{j=1,\, j \neq 1}^{q} x_i^2 \right)$$

y con tres componentes,

$$x_1^2 = x_1 (1 - x_2 - x_3); \quad x_2^2 = x_2 (1 - x_1 - x_3) \quad \text{y} \quad x_3^2 = x_3 (1 - x_1 - x_2)$$

Al aplicar esto en (15.4) y multiplicando a β_0 por $x_1 + x_2 + x_3 = 1$, obtenemos:

$$E(y) = \beta_0\left(x_1 + x_2 + x_3\right) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{11} x_1 (1 - x_2 - x_3) + \beta_{22} x_2 (1 - x_1 - x_3) + \beta_{33} x_3 (1 - x_1 - x_2)$$

multiplicando y reagrupando términos es fácil ver que esto es igual a:

$$E(y) = \beta_1^* x_1 + \beta_2^* x_2 + \beta_3^* x_3 + \beta_{12}^* x_1 x_2 + \beta_{13}^* x_1 x_3 + \beta_{23}^* x_2 x_3 \qquad \textbf{(15.5)}$$

con,

$$\beta_i^* = \beta_0 + \beta_i + \beta_{ij} \quad \text{y} \quad \beta_{ij}^* = \beta_{ij} - \beta_{ii} - \beta_{jj}; \quad i, j = 1, 2, 3$$

En la práctica, para simplificar la notación se eliminan los asteriscos de los coeficientes del modelo (15.5), y de esta manera el modelo canónico de segundo orden para mezclas con tres componentes toma la siguiente forma:

$$E(y) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$

En general, para q componentes este modelo está dado por:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j} \sum_{j=2}^{q} \beta_{ij} x_i x_j$$

donde el coeficiente β_i representa la respuesta esperada en la mezcla pura $x_i = 1$, y al mismo tiempo es la altura de la superficie en el vértice $x_i = 1$. Cuando la mezcla es estrictamente aditiva el polinomio lineal es adecuado (modelo 15.3). Una mezcla es considerada estrictamente aditiva cuando en el centro de las aristas correspondientes a cada par de vértices se predice el promedio de lo estimado en los vértices. Es precisamente el coeficiente β_{ij} el que representa el exceso de la respuesta del modelo cuadrático sobre el lineal, y dependiendo de su signo se habla de *sinergismo* o *antagonismo entre los componentes* correspondientes debido a la mezcla no aditiva. La máxima contribución de este término se da en el punto $x_i = x_j = 1/2$.

En ocasiones, cuando el modelo cuadrático no es suficiente para describir la respuesta puede ajustarse el *modelo cúbico especial*, que para tres componentes está dado por:

$$E(y) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3$$

Mientras que el cúbico completo es,

$$E(y) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \delta_{12} x_1 x_2 (x_1 - x_2) + \delta_{13} x_1 x_3 (x_1 - x_3) + \delta_{23} x_2 x_3 (x_2 - x_3)$$

En general, el modelo cúbico especial para q componentes está dado por:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j} \sum_{j=2}^{q} \beta_{ij} x_i x_j + \sum_{i<j} \sum_{j<k} \sum_{k=3}^{q} \beta_{ijk} x_i x_j x_k \qquad \textbf{(15.6)}$$

mientras que el cúbico completo por:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j}\sum_{j=2}^{q} \beta_{ij} x_i x_j + \sum_{i<j}\sum_{j=2}^{q} \delta_{ij} x_i x_j (x_i - x_j) + \sum_{i<j}\sum_{j<k}\sum_{k=3}^{q} \beta_{ijk} x_i x_j x_k \quad \textbf{(15.7)}$$

En el modelo cúbico, los términos tales como $x_i\, x_j(x_i - x_j)$ permiten mezclas tanto con sinergia como con antagonismo a lo largo del lado $x_i - x_j$. El término cúbico especial $x_1x_2x_3$ cuantifica el efecto de la mezcla ternaria en el interior del simplex.

La interpretación de la magnitud de los coeficientes de un modelo de mezclas debe tratarse en forma especial debido a que $0 \leq x_i \leq 1$. Por ejemplo, un término lineal del tipo β_i únicamente contribuye al modelo cuando $x_i > 0$, y la máxima contribución ocurre cuando $x_i = 1$, en cuyo caso el efecto máximo es igual a β_i. En tanto, un término cuadrático del tipo $\beta_{12}x_1x_2$ contribuye al modelo en cada punto en el simplex donde $x_1 > 0$ y $x_2 > 0$. Su mayor contribución se lleva a cabo en el lado de x_1 y x_2, es de magnitud $(¼)\beta_{12}$ y ocurre en el punto $x_1 = x_2 = ½$. Un término cúbico tal como $\beta_{123}x_1x_2x_3$ contribuye al modelo en cada punto, en el cual $x_1 > 0$, $x_2 > 0$ y $x_3 > 0$ (en el interior del simplex); la máxima contribución es de magnitud $\beta_{123}/27$ y se da en el punto $x_1 = x_2 = x_3 = 1/3$. De lo dicho se desprende que la magnitud de los coeficientes en relación a su contribución sólo es comparable entre los coeficientes de su misma clase: lineales con lineales, cuadráticos con cuadráticos, etcétera.

Ajuste del modelo y caracterización de la superficie de respuesta

Un vez que se cuenta con los datos de un diseño es posible ajustar un modelo sea de primer o mayor grado, según el tipo de diseño. Por lo general, la estimación de los coeficientes del modelo se hace por mínimos cuadrados, como se explicó en el capítulo 11. Para su obtención y para la *caracterización de la superficie de respuesta* es mejor apoyarse en un software estadístico. A continuación se utilizará *Statgraphics* para analizar con detalle los datos de la tabla 15.1 de análogos de queso para la variable y_1.

Para analizar los resultados del diseño de mezcla la primera tarea es ajustar y seleccionar el modelo adecuado. Por ejemplo, en el caso del experimento con análogos de queso ya vimos que se corrió un diseño simplex reticular $\{3, 2\}$ aumentado, por lo que como máximo se puede ajustar un modelo cúbico especial, como el (15.6). Por lo tanto, es preciso decidir si este modelo es adecuado o si es suficiente el modelo cuadrático o incluso el lineal. En la tabla 15.2 se muestran los aspectos básicos para los cuatro posibles modelos.

El *modelo media* consiste sólo en una constante, el *lineal* contiene términos de primer orden para cada x_i. El *cuadrático* incluye las interacciones x_ix_j, y el *cúbico especial* agrega el término $x_1x_2x_3$. El valor-*p* (o *p-value*) prueba si el modelo es significativo o aporta elementos en la explicación de la respuesta al compararlo con el modelo anterior de más bajo orden.

Con el criterio del valor-*p* normalmente se elige el modelo más complicado con un valor-*p* menor que 0.05. Con este criterio se seleccionaría el modelo lineal pero,

como se puede apreciar, el valor-p de los otros modelos también es pequeño (prácticamente menor que 0.10), es decir, también podrían considerarse.

De manera adicional, para la selección del modelo se tiene el *criterio de los coeficientes de determinación* (R^2 y R^2(ajus)), los cuales muestran el porcentaje de la variación en y_1 que es explicada por el correspondiente modelo. En este caso, el *modelo cuadrático* es el indicado, ya que sus coeficientes de determinación son bastante más grandes que los del modelo lineal. Mientras que el *modelo cúbico especial*, aunque podría ser una alternativa, logra aumentar poco el coeficiente de determinación y no compensa la complicación adicional del modelo.

En la tabla 15.3 se muestra el análisis de varianza para el *modelo cuadrático completo*, junto con la estimación de sus coeficientes. El valor-p = 0.0091 en el ANOVA muestra que el modelo es significativo y, como ya habíamos visto, con un R^2 muy bueno. El valor-p = 0.0091 de la tabla 15.3 no coincide con el correspondiente de la tabla 15.2 (p = 0.1016), porque este último indica el aporte adicional del modelo cuadrático respecto al lineal. Como el valor-p para el término x_1x_3 es grande entonces, al excluirlo, se espera que el modelo mejore en términos del R^2(ajus). En efecto, si se elimina este término se observa que el valor-p para el modelo disminuye hasta 0.0016 y la explicación real del modelo dada por R^2(ajus) se incrementa hasta 91.51%. De acuerdo con esto, los coeficientes del modelo cuadrático son ligeramente diferentes a los de la tabla 15.3, y están dados por:

$$y_1 = 0.3208x_1 + 0.7053x_2 + 0.1926x_3 - 0.5338x_1x_2 - 0.5502x_2x_3 \qquad \textbf{(15.8)}$$

en donde se excluye el término x_1x_3, que es claramente no significativo. De la magnitud de los coeficientes lineales estimados se concluye que con la mezcla pura que sólo contiene el componente x_2 (grasa de soya) se logra el valor más grande de y_1 (dureza del análogo de queso) comparado con las otras dos mezclas puras. La mezcla pura que produce los menores valores de y_1 está compuesta sólo con aceite de soya (x_3). En cuanto a los efectos combinados se aprecia que x_2x_3 y x_1x_2 tienen un efecto

Tabla 15.2 Análisis de los posibles modelos para el ejemplo de análogos de queso.

Fuente	Suma de cuadrados	*Gl*	Cuadrado medio	Razón *F*	Valor-*p*
Media	1.156	1	1.156		
Lineal	0.1635	2	0.0817	14.63	0.0032
Cuadrático	0.0296	3	0.0099	4.14	0.1016
Cúbico especial	0.0064	1	0.0064	6.30	0.0870
Error	0.0031	3	0.00102		
Total	1.3586	10			

Modelo	*ES*	R^2	R^2(ajus)
Lineal	0.0747527	80.69	75.18
Cuadrático	0.048791	95.30	89.42
Cúbico especial	0.0320006	98.48	95.45

Tabla 15.3 Ajuste y análisis de varianza para el modelo cuadrático, ejemplo de análogos de queso.

Fuente	Suma de cuadrados	*Gl*	Cuadrado medio	Razón *F*	Valor-*p*
Modelo cuadrático	0.193121	5	0.0386241	16.30	0.0091
Error total	0.00947944	4	0.00236986		
Total (corr.)	0.2026	9			
R^2 = 95.32					
R^2 (ajus) = 89.47					
Error estándar de estimación = 0.04868					
Media del error absoluto = 0.02511					

Parámetro	Estimación	Error estándar	Estadístico *T*	Valor-*p*
x_1	0.3174	0.04696		
x_2	0.7056	0.04696		
x_3	0.1893	0.04696		
x_1x_2	–0.5334	0.21636	–2.465	0.0693
x_1x_3	0.0337	0.21636	0.1557	0.8838
x_2x_3	–0.550	0.21636	–2.541	0.0639

antagónico sobre y_1, dado que el signo de los coeficientes es negativo. De esta manera, si se quisiera maximizar la dureza del análogo del queso, se recomienda formar la fracción grasa sólo con x_2. Pero si se desea otro valor para y_1 habría que explorar la superficie de respuesta del modelo ajustado. En la figura 15.4 precisamente se muestra la gráfica de la superficie de respuesta. Ahí se aprecia claramente que las durezas más bajas se logran con las mezclas puras de x_3 o con mezclas donde este componente predomina. Si se buscara una dureza objetivo de 0.5, se alcanzaría con x_2 próximo a 0.75, y otro 25% se podría completar en forma muy variada con los otros componentes. Obviamente, como en la práctica se quieren otras propiedades en el queso, al final la mezcla ideal se seleccionaría mediante una optimización simultánea. Para ello es necesario ajustar un modelo para cada variable de respuesta y aplicar el procedimiento descrito en el capítulo 13. De manera adicional, el análisis de residuales no detecta ninguna violación a los supuestos del ANOVA.

Interpretación de los coeficientes del modelo ajustado

Coeficiente lineal β_i
No mide el efecto global del ingrediente x_i, sino que sólo estima el valor de la respuesta en ese vértice del simplex.

En diseño de mezclas no es posible interpretar al coeficiente lineal del modelo ajustado como el efecto individual del componente correspondiente. Debido a la restricción de que la suma de las proporciones siempre es igual a la unidad, el efecto de un ingrediente en un diseño de mezclas es diferente al efecto de un factor en un experimento factorial donde se puede mover el factor manteniendo fijos a los demás. En un experimento con mezclas el *coeficiente lineal* β_i no mide el efecto global del ingre-

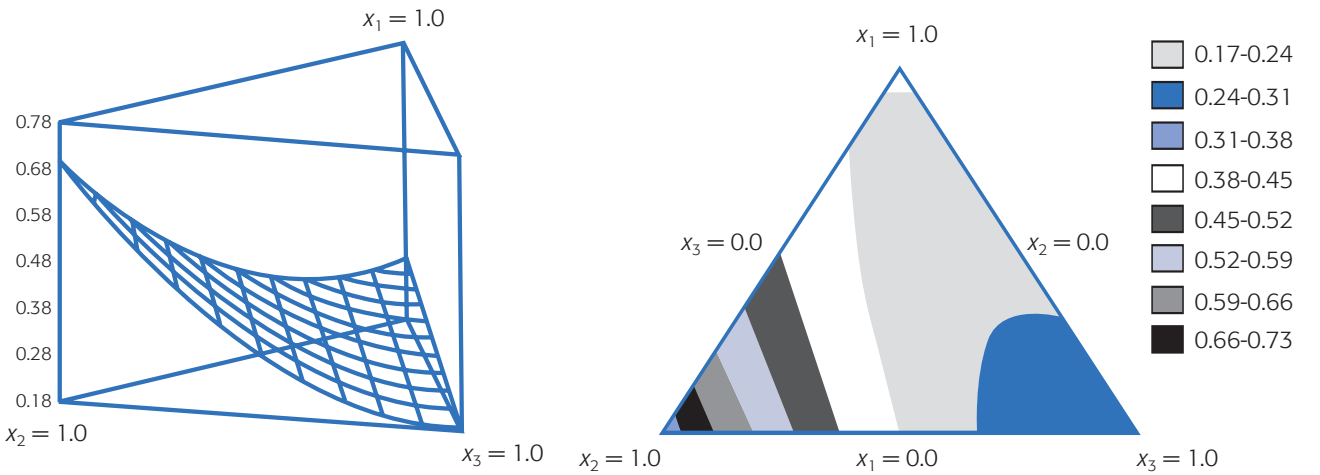


Figura 15.4 Superficie de respuesta para y_1 en análogos de queso.

diente x_i, sino que sólo estima el valor de la respuesta en ese vértice del simplex. Como se observa en la tabla 15.3 no tiene sentido probar las hipótesis H_0: $\beta_i = 0$, puesto que aun cuando el parámetro fuera igual a cero eso no implica que el ingrediente x_i no tenga efecto individual o lineal. Por otra parte, las hipótesis H_0: $\beta_{ij} = 0$ sobre las interacciones sí proveen información acerca del efecto combinado de los componentes x_i y x_j.

El efecto de un ingrediente x_i se define como el cambio en la respuesta a lo largo del eje que une este vértice con el centro del lado opuesto del simplex (véase figura 15.3). Note que a lo largo de este eje la proporción con la que participa el ingrediente x_i cambia de 1 a 0, mientras que el resto de los ingredientes se mantienen en proporciones iguales, cuya suma cambia de 0 a 1.

Gráfico de trazas

Gráfico de trazas
Evalúa la importancia de los componentes de una mezcla. Inicia con una mezcla de referencia y la gráfica va mostrando la manera en que *Y* cambia conforme cada componente aumenta o disminuye su participación en tal mezcla.

El llamado *gráfico de traza* (trace) es una forma de evaluar la importancia de los distintos componentes de una mezcla. Esta gráfica inicia con una mezcla de referencia (usualmente con el centroide de la región experimental), y va mostrando la manera en que la respuesta (Y) se modifica conforme uno de los componentes aumenta o disminuye su participación en la mezcla. Cuando uno de los componentes cambia, el resto se incrementa o disminuye en forma proporcional. De esta manera, la forma del gráfico de trazas muestra la importancia relativa de los componentes.

En la figura 15.5 se muestra el gráfico de trazas para el ejemplo de análogos de queso usando el modelo ajustado dado en la ecuación (15.8). Se observa que el componente con mayor efecto es el x_2, ya que cuando éste es modificado se logran los mayores cambios en y_1. Enseguida se ubicaría al componente x_3 y con menor efecto el x_1. Observe que, de acuerdo a este gráfico, aun cuando el coeficiente de x_3 es menor que el coeficiente de x_1 en la ecuación (15.8), x_3 tiene más efecto que x_1. También se puede ver que el valor máximo de la dureza del queso se alcanza con la mezcla pura $x_2 = 1$, mientras que el valor mínimo se logra en la mezcla pura $x_3 = 1$.

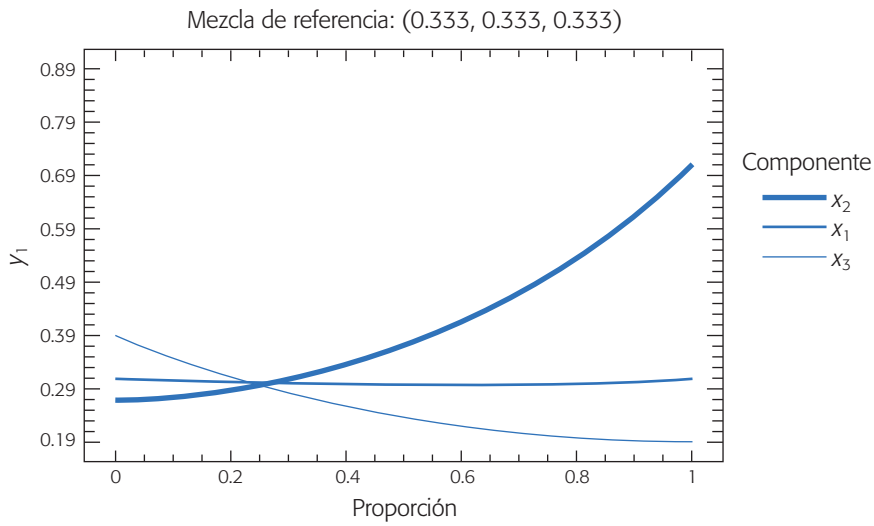


Figura 15.5 Gráfico de trazas para el ejemplo de análogos de queso.

Restricciones en los componentes de una mezcla

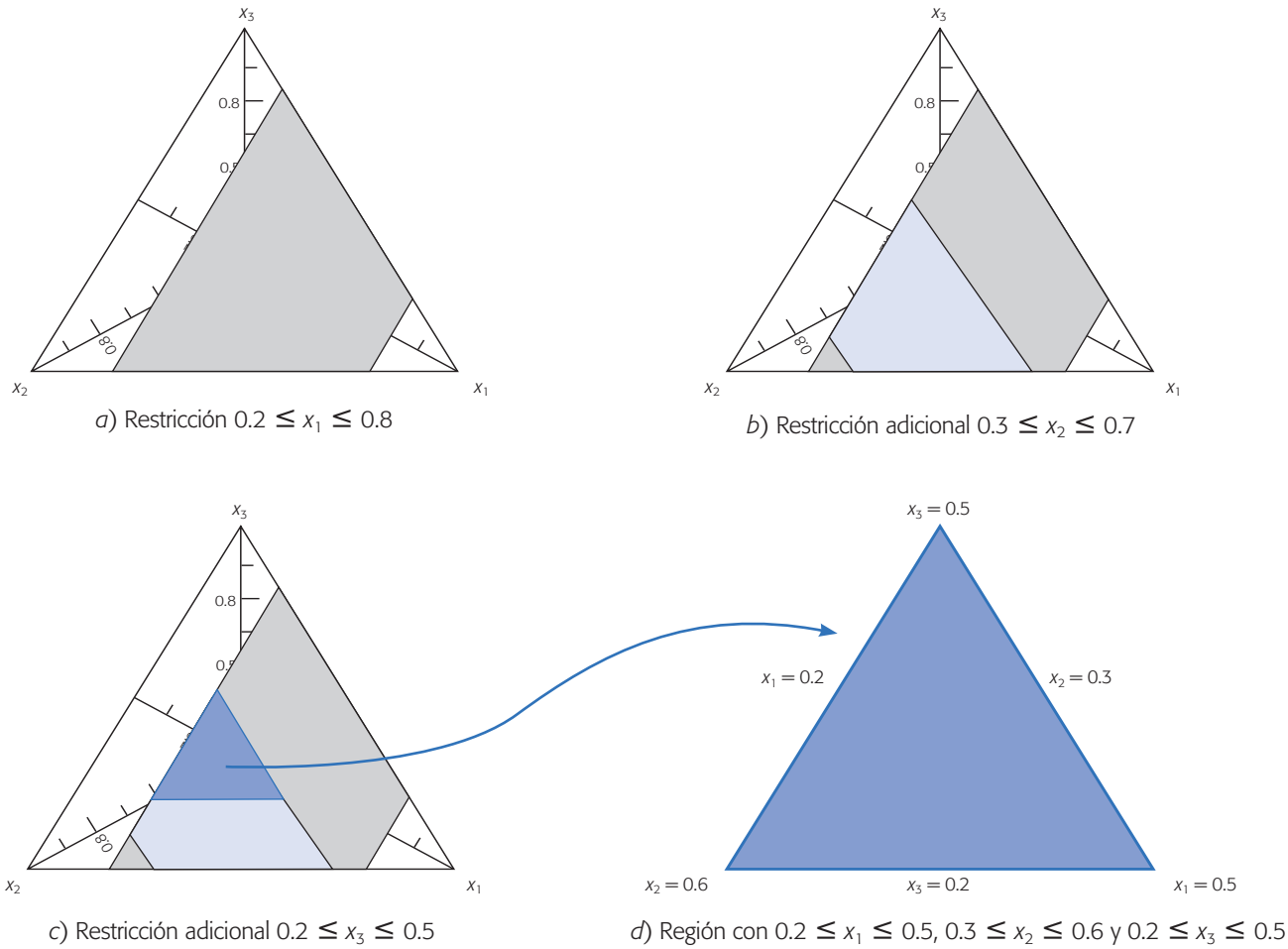
Ya sea por razones técnicas o económicas, es común que algunos componentes en un experimento de mezclas se restrinjan de la siguiente manera:

$$0 \leq a_i \leq x_i \leq b_i \leq 1$$

donde a_i es la restricción inferior para el componente x_i de la mezcla, y b_i es la restricción superior. En esencia, a_i representa la mínima proporción del i-ésimo componente que debe existir en cualquiera de las mezclas que se probarán en el experimento, y b_i representa el nivel o proporción máxima del i-ésimo componente.

En general, las restricciones delimitarán la región experimental factible a una zona o subregión del simplex. Por ejemplo, supongamos que en un experimento con tres componentes el componente x_1 está limitado por $0.2 \leq x_1 \leq 0.8$. Si sólo este componente tuviera restricción, las mezclas a probar tendrían que ser elegidas de la región experimental resaltada en la figura 15.6*a*). De entrada se aprecia que ya no será posible probar ninguna de las mezclas puras. Es decir, al restringir un valor mínimo y máximo a x_1, como se debe cumplir que $x_1 + x_2 + x_3 = 1$, entonces implícitamente se imponen restricciones a los otros componentes.

Si de manera adicional se estableciera la restricción $0.3 \leq x_2 \leq 0.7$, la región factible se muestra en la figura 15.6*b*), donde se observa que esto obliga a que el valor máximo de x_1 ahora sea 0.7, es decir, $0.2 \leq x_1 \leq 0.7$, y en forma automática tendríamos que $0 \leq x_3 \leq 0.5$. Pero si también se deseara que $0.2 \leq x_3 \leq 0.5$, entonces la región experimental factible se muestra en la figura 15.6*c*). Esta última restricción ocasiona que se restrinjan aún más los posibles valores de x_1 y x_2. Se observa que la región experimental final está delimitada por $0.2 \leq x_1 \leq 0.5$, $0.3 \leq x_2 \leq 0.6$ y $0.2 \leq x_3 \leq 0.5$, como se ilustra en la figura 15.6*d*).



a) Restricción $0.2 \leq x_1 \leq 0.8$

b) Restricción adicional $0.3 \leq x_2 \leq 0.7$

c) Restricción adicional $0.2 \leq x_3 \leq 0.5$

d) Región con $0.2 \leq x_1 \leq 0.5$, $0.3 \leq x_2 \leq 0.6$ y $0.2 \leq x_3 \leq 0.5$

Figura 15.6 Regiones experimentales con restricciones en $q = 3$ componentes.

Aunque generalmente no ocurre así, en este caso la forma de la región experimental resultante en la figura 15.6*d*) tiene una forma regular parecida al simplex completo; por ello, para explorar la variable de respuesta en esta región se podrían elegir las mezclas de acuerdo con alguno de los diseños antes descritos: simplex reticular, simplex reticular aumentado o simplex con centroide. Las coordenadas se pueden encontrar por medio de una reparametrización, ver por ejemplo Cornell (2002). Obviamente es más fácil recurrir a un software estadístico. Por ejemplo, los vértices de la región experimental de la figura 15.6*d*) corresponden a las mezclas:

$$(x_1, x_2, x_3) = (0.5, 0.3, 0.2);\ (0.2, 0.6, 0.2) \text{ y } (0.2, 0.3, 0.5)$$

Las coordenadas del centroide están dadas por:

$$(x_1, x_2, x_3) = (0.3, 0.4, 0.3)$$

Cuando la región experimental factible no tiene la forma del simplex, resulta un polígono o hiperpolígono irregular. En esos casos los diseños del tipo simplex no

Diseño de vértices extremos
Se aplica en experimentos de mezclas con restricciones donde la región experimental es un polígono irregular. Se usan los vértices de la región restringida como la base del diseño, junto con puntos entre el centroide global y el centroide de la región experimental.

pueden ser usados, por lo que es necesario recurrir a otros criterios considerados usualmente en los sistemas computacionales. Por ejemplo, un diseño usado con frecuencia en los experimentos de mezclas con restricciones son los que se conocen como vértices extremos propuestos por McLean y Anderson (1966). Ellos sugieren usar los vértices de la región restringida como la base del diseño, junto con una subserie de puntos a lo largo de los centroides de la subregión y el centroide global. También es posible utilizar los criterios *D*-óptimo para seleccionar los puntos. Este criterio fundamentalmente selecciona puntos de una lista de puntos candidatos, de manera tal que las varianzas de los coeficientes de regresión en el modelo con mezclas son minimizados. La efectividad de este método depende de la lista de los puntos candidatos, de la adecuación del modelo propuesto y del número de puntos de diseño seleccionados. Véase Cornell (2002) para mayores detalles.

Los *diseños basados en distancia* también son muy útiles para experimentos con restricciones. Este criterio trata de distribuir de manera uniforme los puntos de diseño en la frontera de la región factible. El algoritmo para seleccionar los puntos inicia con el punto de la región experimental restringida que está más cerca de un vértice de la región sin restricciones, y luego se agrega el punto de la región factible para el cual la distancia euclidiana es un máximo. Todos los puntos subsecuentes son agregados de forma similar.

Ejemplo 15.2

Experimento con restricciones. En García y Espinoza (2006) se describe un experimento en el cual se evalúan los componentes que inciden en el comportamiento reológico de una suspensión de esmalte para baldosas, mediante el uso de experimentos de mezcla con espacio de mezcla restringido. Un esmalte es un revestimiento cerámico constituido por finísimas capas vidriosas que recubren la superficie de la pieza (baldosa en este caso) normalmente con 0.15 a 0.5 mm de espesor.

El proceso de preparación de los esmaltes normalmente consiste en someter la materia prima a una fase de molienda hasta obtener la finura deseada. La materia prima son arcillas, caolines, feldespatos, fritas, colorantes y/o pigmentos, agua y aditivos. Se sabe que los componentes que más influyen son los denominados aditivos. En el proceso del estudio, los aditivos utilizados en la formulación del esmalte fueron: x_1: tripolifosfato (defloculante), x_2: carboximetilcelulosa (ligante o aglutinante) y x_3: sal (suspensivante).

La investigación se enfocó en optimizar la composición de la mezcla con respecto a estos aditivos. Se estableció que la cantidad total de aditivo, resultado de la suma de estos tres componentes, representara 0.245% en el peso de la fórmula total, y que el resto de los componentes del esmalte permanecieran constantes. Para formar este 0.245% asignado a los tres aditivos se decidió aplicar un experimento de mezclas con restricciones, ya que hay razones técnicas que indican el nivel aproximado en que debe estar cada uno de los tres componentes. En la tabla 15.4 se muestran las razones técnicas de dichas restricciones. El diseño experimental de mezclas utilizado fue del tipo *D*-óptimo completamente aleatorizado, el cual se genera a partir de las restricciones usando un software adecuado. Las variables de respuesta fueron y_1: viscosidad, expresada en segundos, y y_2: residuo, expresado en porcentaje. En la tabla 15.5 se muestran el diseño y los resultados obtenidos.

Tabla 15.4 Elección de las restricciones en los componentes del aditivo. El valor entre paréntesis muestra la proporción de cada componente con respecto a 0.245% de aditivo.

Tipo de aditivo	Restricciones (inferior/superior)	Motivo
Componente 1 Trípoli fosfato	0.0375 (0.153)	Un valor inferior aumenta la viscosidad y crea problemas por falta de defloculación.
	0.15 (0.612)	Un valor superior crea problemas de sedimentación, un efecto colateral ocasionado por la reducción de la viscosidad de este aditivo defloculante.
Componente 2 Carboximetilcelulosa	0.075 (0.306)	Un valor inferior crea problemas al momento de la aplicación; por ejemplo, el desconchado, descuelgue y sedimentación en el molino.
	0.1875 (0.765)	Un valor superior aumenta la viscosidad.
Componente 3 Sal de uso industrial	0.013 (0.053)	Valor sugerido según la de conductividad del agua a usar. Conductividad baja implica menor presencia de sales.
	0.0266 (0.109)	Un valor superior crea problemas reológicos pues hay mayor presencia de sales disueltas (mayor conductividad) en el agua a utilizar.

Tabla 15.5 Datos de experimento de mezclas para esmalte.

Orden de corrida	Mezcla	x_1	x_2	x_3	y_1	y_2
19	1	0.184	0.706	0.109	300.00	5.10
12	2	0.384	0.535	0.082	165.63	4.32
11	3	0.335	0.555	0.109	176.27	2.88
6	4	0.155	0.763	0.082	300.00	8.43
15	5	0.384	0.535	0.082	126.18	4.34
1	6	0.392	0.522	0.086	90.68	3.62
16	7	0.249	0.698	0.053	196.66	7.05
18	8	0.612	0.322	0.065	61.39	5.73
9	9	0.527	0.363	0.109	87.06	5.13
3	10	0.612	0.322	0.065	54.77	4.44
5	11	0.584	0.306	0.109	75.87	5.82
20	12	0.584	0.306	0.109	77.18	5.88
2	13	0.184	0.706	0.109	300.00	5.74
13	14	0.384	0.535	0.082	155.38	5.89
7	15	0.506	0.429	0.065	96.43	6.32
8	16	0.322	0.612	0.065	217.15	4.68
10	17	0.441	0.449	0.109	128.08	6.39
14	18	0.384	0.535	0.082	163.59	5.28
17	19	0.155	0.763	0.082	300.00	7.92
4	20	0.249	0.698	0.053	161.84	7.42

En la tabla 15.6 se muestra la información que permite seleccionar el modelo. Con base en el valor-p y R^2 se elige el modelo lineal, ya que explica en forma conveniente la superficie. Ya que la explicación adicional que provee el modelo cuadrático no compensa el tener que agregar los tres términos cuadráticos. Si bien, el término x_1x_2 contribuye de manera significativa (véase tabla 15.7), es posible mostrar que su importancia desaparece si se analiza el logaritmo de la respuesta ($\ln(y_1)$) el cual es la transformación que en este caso mejora el cumplimiento del supuesto de homogeneidad de varianzas (ejercicio 19). En la tabla 15.8 se expone el análisis de varianza y los coeficientes del modelo lineal ajustado:

$$y = 9.34x_1 + 260.4x_2 + 545.6x_3$$

De este modelo se aprecia que el componente x_3, de manera individual y visto como mezcla pura, es el que generaría los valores más grandes de la respuesta y_1 (viscosidad), seguido por x_2 y en tercer lugar x_1. No obstante, debido a las restricciones utilizadas las mezclas puras no se encuentran en la región experimental de interés, por lo que es necesario explorar la superficie de respuesta sobre la región restringida. Los valores deseados para la viscosidad son entre 50 y 60. En la figura 15.7 se muestra la superficie de respuesta en la región experimental, así como la zona óptima delimitada por las curvas de nivel 50 y 60. Cualquier mezcla en esa pequeña zona predice valores de viscosidad en el rango deseado. Por ejemplo, si se quisiera tener $y_1 = 55$, una mezcla óptima sería: $(x_1, x_2, x_3) = (0.60, 0.31, 0.09)$.

El análisis de residuales no muestra problemas graves. Sin embargo, se aprecia un pequeño problema en la gráfica de residuos contra estimados, ya que la varianza de los residuos tiende a incrementarse con los valores de la respuesta. Esto se puede mejorar con la transformación logaritmo de la respuesta. El análisis de la otra variable de respuesta se deja como ejercicio para el lector, junto con la optimización simultánea (véase ejercicio 16).

Tabla 15.6 Análisis de los posibles modelos para el ejemplo de esmalte.

Fuente	Suma de cuadrados	*GI*	Cuadrado medio	Razón *F*	Valor-*p*
Media	522 990.0	1	522 990.0		
Lineal	119 964.0	2	59 982.0	70.88	0.0000
Cuadrático	6 248.82	3	2 082.94	3.58	0.0413<
Cúbico especial	1 005.78	1	1 005.78	1.83	0.1988
Cúbico	3 113.17	3	1 037.72	2.58	0.1116
Error	4 017.57	10	401.757		
Total	657 339.0	20			

Modelo	*ES*	R^2	R^2(ajus)
Lineal	29.0895	89.29	88.03
Cuadrático	24.1077	93.94	91.78
Cúbico especial	23.4205	94.69	92.24
Cúbico	20.0439	97.01	94.32

Tabla 15.7 Modelo cuadrático completo para ejemplo de esmalte.

Parámetro	Estimación	Error estándar	Estadístico *T*	Valor-*p*
x_1	53.4725	25.6603		
x_2	260.053	22.2645		
x_3	–1 418.32	3 585.7		
x_1x_2	–168.512	65.683	–2.56553	0.0224
x_1x_3	1 837.15	4 184.5	0.439038	0.6673
x_2x_3	2 530.3	4 041.89	0.626018	0.5414

Seudocomponentes

Seudocomponentes
Se utilizan para reescalar las proporciones en experimentos de mezclas con restricciones. Facilitan la construcción del diseño y ayudan a hacer el análisis.

Cuando se tienen restricciones en los componentes de la mezcla es recomendable reescalar las proporciones mediante seudocomponentes. Los *seudocomponentes* en experimentos de mezclas son como las variables codificadas en diseños factoriales. En ocasiones, facilitan la construcción del diseño y ayudan a disminuir problemas de multicolinealidad entre las proporciones, que son frecuentes en los diseños de mezclas con restricciones. Si los componentes originales se denotan con x_i ($i = 1, 2,..., q$), los seudocomponentes z_i se definen con la transformación:

$$z_i = \frac{x_i - a_i}{1 - \sum_{i=1}^{q} a_i}; \quad i = 1, 2,..., q$$

donde a_i es la cota inferior del componente x_i. Por ejemplo, cuando sólo existen restricciones inferiores en todos los componentes, la subregión resultante siempre es un simplex y, en este caso, los seudocomponentes convierten la escala de la subregión

Tabla 15.8 Análisis de varianza del modelo lineal, ejemplo de esmalte.

Fuente	Suma de cuadrados	*Gl*	Cuadrado medio	Razón *F*	Valor-*p*
Modelo lineal	119 959.0	2	59 979.7	70.86	0.0000
Error total	14 390.1	17	846.475		
Total (corr.)	134 349.0	19			
$R^2 = 89.29$					
R^2 (ajus) = 88.03					
Error estándar de estimación = 29.1					
Media del error absoluto = 21.86					

Parámetro	Estimación	Error estándar	Estadístico *T*	Valor-*p*
x_1	9.34	17.268		
x_2	260.457	15.593		
x_3	545.636	151.897		

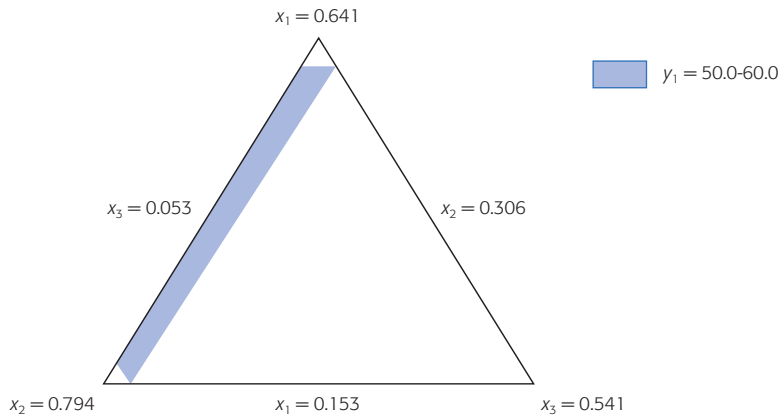


Figura 15.7 Región experimental y zona óptima para y_1.

en la escala del simplex original, sobre el cual se definen directamente los diseños de mezclas usuales. Pero aun cuando la subregión no sea un simplex, siempre es recomendable utilizar seudocomponentes. De hecho algunos software, como *Statgraphics*, los usan de manera automática para estimar el modelo ajustado.

La idea de diseños mixtos: componentes de mezclas y variables de proceso

Como parte esencial de este capítulo es importante destacar que en algunos problemas de mezclas, además del problema de estudiar el efecto de los componentes de la mezcla es de interés investigar la influencia de variables de proceso. Éstas se refieren a los factores que afectan la respuesta estudiada pero que no corresponden a proporciones de los componentes en las mezclas; son el tipo de factores que se presentan en el diseño de experimentos clásico, como temperaturas, tiempos, etc. Precisamente a través de un *experimento mixto* se estudian los efectos combinados de variables de proceso y componentes de una mezcla sobre una determinada respuesta.

Experimento mixto
Con este diseño se estudian los efectos combinados de variables de proceso y componentes de una mezcla sobre una determinada variable de respuesta.

Los diseños experimentales que comúnmente se utilizan consisten en probar los diferentes tratamientos de un diseño de mezclas en cada una de las combinaciones de los factores de proceso. Por ejemplo, si se tienen tres componentes de mezclas y dos variables de proceso, entonces se podrían probar las siete mezclas de un diseño simplex con centroide en cada una de las cuatro condiciones de proceso de un diseño 2^2. De esta manera, en total se tendrían 28 tratamientos que se corren en orden aleatorio. Al lector interesado en este tema se le recomienda consultar Cornell (2002).

Uso de software estadístico

El software *Statgraphics* también incluye los diseños de mezclas que se explicaron en este capítulo. Se accede a ellos con la siguiente secuencia: *Special* → *Expermental Design* → *Create Design* y *Mixture* (o *Doe* → *Design creation* → *Create New Desing* → *Mixture*). En el análisis se incluyen todos los procedimientos descritos.

En *Minitab* también se incluyen estos diseños, y se accede por medio de la secuencia: *Stat* → *DOE* → *Mixture* → *Create Mixture Design*.

Preguntas y ejercicios

1. Explique en qué situaciones se aplican los experimentos con mezclas.
2. ¿Por qué los diseños factoriales tradicionales, del tipo 2^k, no se aplican en experimentos con mezclas?
3. Explique en qué consiste un diseño simplex reticular y represente uno del tipo {3, 3}.
4. Comience con un diseño simplex reticular {3, 2}, auméntelo y diga por qué razón se hace esto.
5. En qué consiste un diseño simplex con centroide. Ejemplifique para $q = 4$ componentes.
6. Anote los tratamientos (mezclas) que se ejecutan en un diseño simplex reticular {4, 2}.
7. ¿Generalmente cuándo se emplea un diseño de mezclas con restricciones? Explique.
8. Considere diseños de mezclas con $q = 3$ componentes y conteste lo siguiente:
   a) ¿Qué implica que uno de los componentes tenga restricciones, por ejemplo, $1/6 \leq x_1 \leq 4/6$?
   b) Si sólo se tiene la restricción $1/6 \leq x_1 \leq 4/6$, represéntela en forma gráfica y establezca las restricciones implícitas que surgen para los otros dos componentes.
   c) Además de la restricción $1/6 \leq x_1 \leq 4/6$, se tiene que $1/3 \leq x_2 \leq 2/3$. Represente ambas restricciones en forma gráfica. ¿En realidad se pueden cumplir completamente ambas restricciones? Argumente.
   d) En el caso anterior, ¿cuál es la restricción que surge para x_3?
9. Represente en forma gráfica la región experimental para $q = 3$ componentes de una mezcla, partiendo de las siguientes restricciones $1/6 \leq x_1 \leq 4/6$, $1/3 \leq x_2 \leq 2/3$ y $1/3 \leq x_3 \leq 1/2$. Después conteste lo siguiente:
   a) ¿Es factible cumplir con estas restricciones? Explique.
   b) En caso de que no sea posible, proponga restricciones para los tres componentes que sí se puedan cumplir.
10. Con respecto al problema de elaboración de análogos de queso descrito en el ejemplo 1, además de dureza (y_1, kg), elasticidad (y_2, cm) se evaluaron otras variables, entre ellas firmeza sensorial (y_3) y cremosidad (y_4). Las últimas dos variables se evaluaron en forma sensorial mediante un panel de seis jueces, quienes ordenaron las muestras de menor a mayor intensidad y de acuerdo a la característica a evaluar. Con esto se obtuvieron el índice *R*, que representa la probabilidad de que los jueces detecten diferencias entre el tratamiento y el análogo de referencia (el elaborado con grasa butírica). Por lo tanto, un valor cercano a 50 indicará que no hay diferencias (que es lo que se desea). Los resultados promedio con tres réplicas se muestran a continuación:

x_1	x_2	x_3	y_1	y_2	y_3	y_4
1	0	0	0.32	0.97	50	50
0	1	0	0.7	0.83	96.7	0.92
0	0	1	0.2	0.87	0.9	35.1
0.5	0.5	0	0.33	0.7	91.2	39.8
0.5	0	0.5	0.23	0.82	3.12	77
0	0.5	0.5	0.27	0.74	35.6	54.6
0.333	0.333	0.333	0.31	0.78	26.8	66.1
0.667	0.167	0.167	0.32	0.76	21	61.4
0.167	0.667	0.167	0.49	0.73	79.1	37.5
0.167	0.167	0.667	0.23	0.82	1.8	93

a) ¿Qué diseño se utilizó?
b) Haga un análisis completo para cada una de las variables de respuesta.

11. En Vargas del Río *et al*. (2005) se presenta un estudio para evaluar la posibilidad de utilizar puzolana como material alternativo en la construcción. Se aplica un diseño de experimentos de mezclas, cuyos componentes son: puzolana, cal y cemento; y como variable de respuesta se tiene la resistencia a la comprensión de la mezcla. Se decide emplear un diseño simplex reticular (3, 3) con dos réplicas en todos los puntos, y se agregan seis mezclas de particular interés (las últimas seis mezclas) en la siguiente tabla.

Puzolana	Cal	Cemento	Réplica 1	Réplica 2
1	0	0	0	0
0.666	0.333	0	43.91	47.01
0.666	0	0.333333	245.67	264.48
0.333	0.666	0	11.95	11.13
0.333	0.333	0.333	113.67	120.19
0.333	0	0.666	499.19	420.4
0	1	0	5.13	3.08
0	0.666	0.333	35.97	32.69
0	0.333	0.666	269.73	244.79
0	0	1	617.78	599.94
0.666	0.166	0.166	132.67	121.34
0.166	0.666	0.166	13.26	18.4
0.166	0.166	0.666	317.01	328.86
0.75	0.25	0	49.99	50.82
0.85	0.15	0	64.2	62.87
0.15	0	0.85	613.71	613.71

a) Represente gráficamente el diseño empleado.
b) Ajuste el modelo apropiado, anote la ecuación y señale si éste es suficiente para explicar lo que ocurre en el experimento.
c) Por medio de la gráfica de trazas jerarquice en orden de importancia los materiales que tienen mayor impacto sobre la resistencia, y diga si hay alguna sinergia de relevancia o antagonismo. Justifique sus afirmaciones.
d) Bosqueje la gráfica de contornos y señale en qué zona se dan las respuestas máximas.
e) Encuentre la mezcla óptima.

12. En Cornell (2002) se describe un experimento en el que se desea elaborar una mezcla a partir de cuatro pesticidas, Vendex (x_1), Omite (x_2), Kelthane (x_3) y Dibrom (x_4) para controlar la plaga de gorgojos en los cultivos de fresas. La efectividad de cada tratamiento se midió con la respuesta y: porcentaje de gorgojos muertos (al comparar la cantidad antes y después de aplicar el tratamiento en 10 hojas de tres plantas). A continuación se muestra el diseño empleado y los resultados obtenidos.

x_1	x_2	x_3	x_4	y
0	0	0	1	1.8
0	1	0	0	25.4
0	0	1	0	28.6
1	0	0	0	38.5
0	0.5	0	0.5	4.9
0	0	0.5	0.5	3.1
0.5	0	0	0.5	28.7
0	0.5	0.5	0	3.4
0.5	0.5	0	0	37.4
0.5	0	0.5	0	10.7
0	0.33	0.33	0.33	22.0
0.33	0.33	0	0.33	2.4
0.33	0	0.33	0.33	2.6
0.33	0.33	0.33	0	11.1
0.25	0.25	0.25	0.25	0.8

a) ¿Qué diseño se empleó?
b) Ajuste el modelo apropiado y señale si éste es suficiente para explicar lo que ocurre en el experimento.
c) Anote la ecuación del modelo y explique los hechos más sobresalientes que se derivan del mismo.
d) Obtenga la gráfica de trazas e interprétela.
e) Bosqueje la gráfica de contornos con los tres componentes que mejor le ayuden a analizar la superficie, y señale en qué zona se dan las respuestas máximas.
f) Encuentre la mezcla óptima. ¿Hay algún problema con que esta mezcla no se haya corrido?
g) ¿Los residuales muestran algún problema potencial?

13. Anderson y Whitcomb (2002) reportan un experimento con solventes donde el objetivo es estudiar la solubilidad en función de tres ingredientes: MEK (metil-etil-ketano), tolueno y hexano. El experimento expresado en porcentajes, así como los datos observados se muestran en la siguiente tabla:

MEK	Tolueno	Hexano	Solubilidad (g/L)
100	0	0	121
0	100	0	164
0	0	100	179
50	50	0	140
0	50	50	180
50	0	50	185
33.3	33.3	33.3	199
66.6	16.7	16.7	175
16.7	66.6	16.7	186
16.7	16.7	66.6	201

a) Ajuste el modelo canónico más adecuado para estos datos.
b) ¿Es posible afirmar que no existen efectos antagónicos? Argumente.
c) Determine cuál de los ingredientes tiene mayor efecto sobre la solubilidad. Utilice la gráfica de traza.
d) Suponga que interesa maximizar la solubilidad y dibuje la curva de nivel que corresponde a una solubilidad igual a 204. Use la gráfica para determinar la mezcla óptima.
e) Verifique los supuestos del modelo.

14. Frisbee y McGinity (1994) presentan un experimento para estudiar el efecto de tres agentes de superficie activa sobre la dispersión acuosa de nanoesferas poliméricas. También estudiaron las propiedades de formado del velo de este producto farmacéutico. El experimento y las dos respuestas medidas se muestran en la siguiente tabla:

A	*B*	*C*	Tamaño de partícula	Temp. de transición
1	0	0	250.1	18.9
0	1	0	274.1	15.2
0	0	1	533.5	35.0
0.5	0.5	0	255.2	16.1
0.5	0	0.5	267.3	18.9
0	0	0.5	294.3	31.2
0.333	0.333	0.333	250.5	19.3
0.666	0.167	0.167	232.5	18.2
0.167	0.666	0.167	251.0	17.7
0.167	0.167	0.666	276.0	30.1
0.333	0.333	0.333	255.0	19.0

a) Ajuste un modelo apropiado para cada respuesta.
b) Dibuje las dos superficies utilizando gráficas de contornos.
c) Ambas respuestas se quieren minimizar. Encuentre las mezclas óptimas individuales y la mezcla óptima global, que es la que finalmente interesa.

15. En Espinoza Escalante *et al*. (2006) se presenta un experimento para encontrar una mezcla óptima de vinaza, inóculo y agua; para maximizar la producción de biomasa se corrió un diseño simplex-centroide con restricciones. La mezcla se desarrolló a dos temperaturas.

Mezcla	Vinaza	Inóculo	Agua	%Δ de biomasa (35°C)	%Δ de biomasa (55°C)
1	90	0	10	13.4	22.7
2	61.5	28.5	10	12.4	22.6
3	33.0	57.0	10	12.1	5.6
4	52.0	19.0	29.0	7.6	8.8
5	61.5	0	38.5	7.4	12.1
6	33	28.5	38.5	7.1	13.7
7	33	0	67	6.1	3.8

a) ¿Cuáles son las restricciones para cada componente de la mezcla?
b) Bosqueje en un simplex la región de experimentación.
c) Haga un análisis de mezclas en cada temperatura y obtenga conclusiones.

16. Con respecto al problema de elaborar un esmalte variando la composición de los aditivos, descrito en el ejemplo 15.2, realice lo siguiente:
a) Haga un análisis completo para la variable y_2.
b) Proponga una combinación de (x_1, x_2, x_3) si se quiere lograr un valor de y_2 entre 5 y 6.
c) Haga una optimización simultánea con base en la función de deseabilidad (capítulo 13) para ambas variables y_1, y_2.

17. En Bowles y Montgomery (1997) se presenta el problema de elaborar una margarita (bebida con tequila). Diseñar la fórmula de una margarita es un problema que corresponde al ámbito de los experimentos con mezclas, donde los ingredientes tienen restricciones, ya que una mezcla que contenga 100% de cualquiera de los cuatro ingredientes resultará poco satisfactoria y se desea que cada componente esté presente en cualquier formulación. Con base en la experiencia práctica con esta bebida se definieron las siguientes restricciones:

Margarita Mix:	$0.49 \le x_1 \le 0.55$
Tequila:	$0.25 \le x_2 \le 0.31$
Triple Sec:	$0.08 \le x_3 \le 0.16$
Jugo de lima:	$0.04 \le x_4 \le 0.10$

Con $x_1 + x_2 + x_3 + x_4 = 1$. La variable de respuesta de este experimento debe ser una medida del goce de esta bebida. Se usaron dos variables de respuesta, y_2: los panelistas evaluaron cada receta sobre una escala del 1 al 10, indicando que el 10 se refiere a la más deliciosa (sabor y la intensidad de la bebida); y_1: cada participante fue cuestionado para comparar el goce relacionado con cada receta, construyendo una escala en la que la mejor se considerará en el lugar número 1, la siguiente ocupará el número 2 y así sucesivamente. Se formó un jurado con 15 participantes. Se reportan los resultados promedio.

Mezcla	x_1	x_2	x_3	x_4	y_1	y_2
1	49	25	16	10	9.33	5.14
2	51	31	8	10	9.17	4.43
3	55	28	13	4	9.57	3.96
4	55	25	10	10	3.67	6.29
5	55	29	8	8	7.33	5.07
6	51	31	8	10	9.00	4.12
7	55	25	10	10	5.33	6.21
8	52	28	16	4	12.33	4.21
9	52	25	16	7	8.67	5.54
10	49	28	13	10	6.50	4.92
11	55	25	16	4	8.00	4.60
12	55	29	8	8	7.17	5.79
13	55	31	10	4	7.50	4.60
14	49	31	13	7	7.67	4.99
15	49	31	16	4	9.17	2.79

a) Ajustar el modelo apropiado y depurar el modelo para y_2.
b) De acuerdo con el modelo ajustado, ¿cuáles son los aspectos más relevantes?
c) ¿Cuál sería una mezcla óptima?
d) Repita el análisis anterior ahora para y_1.
e) ¿Qué combinación es satisfactoria para ambas variables?

18. En Vargas Marín, *et al*. (2006) se presentan los resultados de un experimento de mezclas con restricciones donde se estudia el efecto de los componentes que forma un tabicón, es decir, x_1: agua, x_2: cemento pórtland, x_3: jal grueso, x_4: jal fino y x_5: arena amarilla. Las variables de respuesta medidas a los tabicones fueron y_1: resistencia a compresión, y_2: contracción por secado, y_3: absorción y y_4: peso específico. Las restricciones se establecieron a partir de conocimientos técnicos sobre distintos tipos de tabicón. El diseño empleado y los resultados experimentales se muestran a continuación:

Mezcla	x_1	x_2	x_3	x_4	x_5	y_1	y_2	y_3	y_4
1	0.029	0.143	0.429	0.200	0.200	81.3	0.211	258.4	1 370.2
2	0.029	0.143	0.200	0.429	0.200	99.9	0.212	252.5	1 443.5
3	0.029	0.143	0.200	0.200	0.429	85.0	0.199	229.3	1 493.1
4	0.075	0.143	0.382	0.200	0.200	150.4	0.238	172.3	1 350.1
5	0.075	0.143	0.200	0.382	0.200	149.2	0.205	207.9	1 348.8
6	0.075	0.143	0.200	0.200	0.382	162.2	0.178	233.5	1 345.0
7	0.029	0.121	0.450	0.200	0.200	125.1	0.291	269.3	1 372.3
8	0.029	0.063	0.450	0.259	0.200	68.4	0.294	286.5	1 363.7
9	0.029	0.063	0.450	0.200	0.259	53.8	0.247	302.1	1 330.8
10	0.075	0.075	0.450	0.200	0.200	89.5	0.277	254.6	1 270.5
11	0.075	0.063	0.450	0.213	0.200	78.4	0.330	302.2	1 199.7
12	0.075	0.063	0.450	0.200	0.213	72.7	0.237	302.6	1 209.3
13	0.029	0.121	0.200	0.450	0.200	111.4	0.250	221.6	1 338.9
14	0.029	0.063	0.259	0.450	0.200	72.9	0.257	266.6	1 170.7
15	0.029	0.063	0.200	0.450	0.259	46.1	0.191	243.5	1 212.4
16	0.075	0.075	0.200	0.450	0.200	89.3	0.250	283.4	1 262.1
17	0.075	0.063	0.213	0.450	0.200	81.5	0.303	271.7	1 225.1
18	0.075	0.063	0.200	0.450	0.213	76.4	0.309	245.9	1 245.5
19	0.029	0.121	0.200	0.200	0.450	112.4	0.251	293.5	1 382.0
20	0.029	0.063	0.259	0.200	0.450	107.0	0.311	288.9	1 405.3
21	0.029	0.063	0.200	0.259	0.450	83.9	0.155	290.3	1 373.4
22	0.075	0.075	0.200	0.200	0.450	110.5	0.120	313.1	1 301.5
23	0.075	0.063	0.213	0.200	0.450	69.5	0.154	319.2	1 286.6
24	0.075	0.063	0.200	0.213	0.450	65.0	0.199	311.1	1 265.6
25	0.050	0.100	0.283	0.283	0.283	100.4	0.199	249.4	1 294.0
26	0.047	0.093	0.300	0.280	0.280	110.2	0.151	250.8	1 299.0
27	0.053	0.093	0.277	0.300	0.277	105.9	0.078	211.5	1 280.0
28	0.050	0.107	0.280	0.283	0.280	119.0	0.078	243.2	1 342.3
29	0.050	0.100	0.277	0.277	0.297	137.4	0.076	278.1	1 288.0

a) ¿Cuáles son las restricciones para cada componente?
b) ¿Por qué son necesarias estas restricciones? ¿Por qué cree que no se probaron las mezclas puras?
c) ¿Es posible representar gráficamente la región experimental?
d) Para cada variable de respuesta ajuste el modelo apropiado, depúrelo, anote la ecuación y señale si éste es suficiente para explicar lo que ocurre en el experimento.
e) Interprete con detalle cada modelo ajustado: jerarquizando por orden de importancia los efectos de mayor impacto sobre cada *y*. Use la gráfica de traza.
f) Considere los componentes con mayor impacto, seleccione tres, después obtenga y analice la correspondiente gráfica de contornos.
g) Considere los siguientes valores deseables para las variables de respuesta: y_1 maximizar ($y_1 > 100$), y_2 minimizar, y_3 minimizar ($y_3 < 290$) y y_4 minimizar; después, encuentre para cada variable la mezcla óptima.

19. Considere los datos del ejemplo 15.2 que se dan en la tabla 15.6.
a) Haga otra vez en análisis del experimento, pero ahora considerando como respuesta el logaritmo natural de la viscosidad.
b) Comente las diferencias observadas entre el análisis dado en el ejemplo y este nuevo análisis. En particular, compare la significancia del término x_1x_2 y el porcentaje de explicación del modelo lineal.
c) También compare la gráfica de residuos contra predichos de ambos análisis.

Apéndice A
Tablas

Sumario

- A1: Tablas de la distribución normal
- A2: Tablas de la distribución χ^2
- A3: Tablas *T* de Student
- A4: Tablas de la distribución *F*
- A5: Tablas de rango estudentizado
- A6: Tablas para comparación de medias Duncan
- A7: Tablas de Dunnet
- A8: Tablas para prueba de independencia Durbin-Watson

Tabla A1 Puntos críticos de la distribución normal estándar ($\mu = 0, \sigma = 1$), $P(Z > z)$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.9
0.0	0.500	0.496	0.492	0.488	0.484	0.480	0.476	0.472	0.468	0.464
0.1	0.460	0.456	0.452	0.448	0.444	0.440	0.436	0.433	0.429	0.425
0.2	0.421	0.417	0.413	0.409	0.405	0.401	0.397	0.394	0.390	0.386
0.3	0.382	0.378	0.374	0.371	0.367	0.363	0.359	0.356	0.352	0.348
0.4	0.345	0.341	0.337	0.334	0.330	0.326	0.323	0.319	0.316	0.312
0.5	0.309	0.305	0.302	0.298	0.295	0.291	0.288	0.284	0.281	0.278
0.6	0.274	0.271	0.268	0.264	0.261	0.258	0.255	0.251	0.248	0.245
0.7	0.242	0.239	0.236	0.233	0.230	0.227	0.224	0.221	0.218	0.215
0.8	0.212	0.209	0.206	0.203	0.200	0.198	0.195	0.192	0.189	0.187
0.9	0.184	0.181	0.179	0.176	0.174	0.171	0.169	0.166	0.164	0.161
1.0	0.159	0.156	0.154	0.152	0.149	0.147	0.145	0.142	0.140	0.138
1.1	0.136	0.133	0.131	0.129	0.127	0.125	0.123	0.121	0.119	0.117
1.2	0.115	0.113	0.111	0.109	0.107	0.106	0.104	0.102	0.100	0.099
1.3	0.097	0.095	0.093	0.092	0.090	0.089	0.087	0.085	0.084	0.082
1.4	0.081	0.079	0.078	0.076	0.075	0.074	0.072	0.071	0.069	0.068
1.5	0.067	0.066	0.064	0.063	0.062	0.061	0.059	0.058	0.057	0.056
1.6	0.055	0.054	0.053	0.052	0.050	0.049	0.048	0.047	0.046	0.046
1.7	0.045	0.044	0.043	0.042	0.041	0.040	0.039	0.038	0.038	0.037
1.8	0.036	0.035	0.034	0.034	0.033	0.032	0.031	0.031	0.030	0.029
1.9	0.029	0.028	0.027	0.027	0.026	0.026	0.025	0.024	0.024	0.023
2.0	0.023	0.022	0.022	0.021	0.021	0.020	0.020	0.019	0.019	0.018
2.1	0.018	0.017	0.017	0.017	0.016	0.016	0.015	0.015	0.015	0.014
2.2	0.014	0.014	0.013	0.013	0.013	0.012	0.012	0.012	0.011	0.011
2.3	0.011	0.010	0.010	0.010	0.010	0.009	0.009	0.009	0.009	0.008
2.4	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.006
2.5	0.006	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.005
2.6	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
2.7	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
2.8	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
2.9	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001
3.0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
3.1	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
3.2	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000
3.3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Tabla A2 Puntos para la distribución χ^2.

	Área a la derecha de los puntos, $P(X > x)$					
Grados de libertad	**0.995**	**0.975**	**0.95**	**0.05**	**0.025**	**0.015**
1	0.000	0.001	0.004	3.841	5.024	5.916
2	0.010	0.051	0.103	5.991	7.378	8.399
3	0.072	0.216	0.352	7.815	9.348	10.465
4	0.207	0.484	0.711	9.488	11.143	12.339
5	0.412	0.831	1.145	11.070	12.833	14.098
6	0.676	1.237	1.635	12.592	14.449	15.777
7	0.989	1.690	2.167	14.067	16.013	17.398
8	1.344	2.180	2.733	15.507	17.535	18.974
9	1.735	2.700	3.325	16.919	19.023	20.513
10	2.156	3.247	3.940	18.307	20.483	22.021
11	2.603	3.816	4.575	19.675	21.920	23.503
12	3.074	4.404	5.226	21.026	23.337	24.963
13	3.565	5.009	5.892	22.362	24.736	26.403
14	4.075	5.629	6.571	23.685	26.119	27.827
15	4.601	6.262	7.261	24.996	27.488	29.235
16	5.142	6.908	7.962	26.296	28.845	30.629
17	5.697	7.564	8.672	27.587	30.191	32.011
18	6.265	8.231	9.390	28.869	31.526	33.382
19	6.844	8.907	10.117	30.144	32.852	34.742
20	7.434	9.591	10.851	31.410	34.170	36.093
21	8.034	10.283	11.591	32.671	35.479	37.434
22	8.643	10.982	12.338	33.924	36.781	38.768
23	9.260	11.689	13.091	35.172	38.076	40.094
24	9.886	12.401	13.848	36.415	39.364	41.413
25	10.520	13.120	14.611	37.652	40.646	42.725
26	11.160	13.844	15.379	38.885	41.923	44.031
27	11.808	14.573	16.151	40.113	43.195	45.331
28	12.461	15.308	16.928	41.337	44.461	46.626
29	13.121	16.047	17.708	42.557	45.722	47.915
30	13.787	16.791	18.493	43.773	46.979	49.199

Tabla A3 Puntos críticos para la distribución *T* de Student.

	Área a la derecha de los puntos, $P(X > x)$			
Grados de libertad	**0.10**	**0.05**	**0.025**	**0.015**
1	3.0776835	6.3137515	12.706205	21.204949
2	1.8856181	2.9199856	4.3026527	5.6427784
3	1.6377444	2.3533634	3.1824463	3.8960459
4	1.5332063	2.1318468	2.7764451	3.2976297
5	1.475884	2.0150484	2.5705818	3.002875
6	1.4397557	1.9431803	2.4469119	2.8289279
7	1.4149239	1.8945786	2.3646243	2.714573
8	1.3968153	1.859548	2.3060041	2.6338144
9	1.3830287	1.8331129	2.2621572	2.573804
10	1.3721836	1.8124611	2.2281389	2.5274842
11	1.3634303	1.7958848	2.2009852	2.4906639
12	1.3562173	1.7822876	2.1788128	2.4607002
13	1.3501713	1.7709334	2.1603687	2.4358452
14	1.3450304	1.7613101	2.1447867	2.4148977
15	1.3406056	1.7530504	2.1314495	2.397005
16	1.3367572	1.7458837	2.1199053	2.3815454
17	1.3333794	1.7396067	2.1098156	2.3680548
18	1.3303909	1.7340636	2.100922	2.35618
19	1.3277282	1.7291328	2.0930241	2.3456475
20	1.3253407	1.7247182	2.0859634	2.3362422
21	1.3231879	1.7207429	2.0796138	2.3277923
22	1.3212367	1.7171444	2.0738731	2.3201596
23	1.3194602	1.7138715	2.0686576	2.313231
24	1.3178359	1.7108821	2.0638986	2.3069134
25	1.3163451	1.7081408	2.0595386	2.3011295
26	1.3149719	1.7056179	2.0555294	2.2958145
27	1.3137029	1.7032884	2.0518305	2.2909136
28	1.3125268	1.7011309	2.0484071	2.2863802
29	1.3114336	1.699127	2.0452296	2.2821746
30	1.310415	1.6972609	2.0422725	2.2782623

Tabla A4 Puntos críticos al 5% de la distribución F, $P(X > x) = 0.05$.

Grados de libertad en el denominador	Grados de libertad en el numerador																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	25	30	40	50	75	100	∞
1	**161**	**199**	**216**	**225**	**230**	**234**	**237**	**239**	**241**	**242**	**243**	**244**	**245**	**245**	**246**	**248**	**249**	**250**	**251**	**252**	**253**	**253**	**254**
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	91.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70	8.66	8.63	8.62	8.59	8.58	8.56	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86	5.80	5.77	5.75	5.72	5.70	5.68	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62	4.56	4.52	4.50	4.46	4.44	4.42	4.41	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94	3.87	3.83	3.81	3.77	3.75	3.73	3.71	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51	3.44	3.40	3.38	3.34	3.32	3.29	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22	3.15	3.11	3.08	3.04	3.02	2.99	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.94	2.89	2.86	2.83	2.80	2.77	2.76	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.77	2.73	2.70	2.66	2.64	2.60	2.59	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.65	2.60	2.57	2.53	2.51	2.47	2.46	2.41
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.54	2.50	2.47	2.43	2.40	2.37	2.35	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.46	2.41	2.38	2.34	2.31	2.28	2.26	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.39	2.34	2.31	2.27	2.24	2.21	2.19	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.33	2.28	2.25	2.20	2.18	2.14	2.12	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.28	2.23	2.19	2.15	2.12	2.09	2.07	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.23	2.18	2.15	2.10	2.08	2.04	2.02	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.19	2.14	2.11	2.06	2.04	2.00	1.98	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.16	2.11	2.07	2.03	2.00	1.96	1.94	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.12	2.07	2.04	1.99	1.97	1.93	1.91	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18	2.10	2.05	2.01	1.96	1.94	1.90	1.88	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15	2.07	2.02	1.98	1.94	1.91	1.87	1.85	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18	2.15	2.13	2.05	2.00	1.96	1.91	1.88	1.84	1.82	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13	2.11	2.03	1.97	1.94	1.89	1.86	1.82	1.80	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09	2.01	1.96	1.92	1.87	1.84	1.80	1.78	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.12	2.09	2.07	1.99	1.94	1.90	1.85	1.82	1.78	1.76	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.10	2.08	2.06	1.97	1.92	1.88	1.84	1.81	1.76	1.74	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.09	2.06	2.04	1.96	1.91	1.87	1.82	1.79	1.75	1.73	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.08	2.05	2.03	1.94	1.89	1.85	1.81	1.77	1.73	1.71	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01	1.93	1.88	1.84	1.79	1.76	1.72	1.70	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92	1.84	1.78	1.74	1.69	1.66	1.61	1.59	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86	1.84	1.75	1.69	1.65	1.59	1.56	1.51	1.48	1.39
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.82	1.79	1.77	1.68	1.62	1.57	1.52	1.48	1.42	1.39	1.28
∞	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.72	1.69	1.67	1.57	1.51	1.46	1.40	1.35	1.28	1.25	1.03

Tabla A4 (*Continuación*) Puntos críticos al 10% de la distribución F, $P(X > x) = 0.10$.

Grados de libertad en el denominador	Grados de libertad en el numerador																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	25	30	40	50	75	100	∞
1	40	50	54	56	57	58	59	59	60	60	60	61	61	61	61	62	62	62	63	63	63	63	63
2	8.5	9.0	9.2	9.2	9.3	9.3	9.3	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.5	9.5	9.5	9.5	9.5	9.5	9.5
3	5.5	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22	5.21	5.20	5.20	5.18	5.17	5.17	5.16	5.15	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90	3.89	3.88	3.87	3.84	3.83	3.82	3.80	3.80	3.78	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27	3.26	3.25	3.24	3.21	3.19	3.17	3.16	3.15	3.13	3.13	3.11
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90	2.89	2.88	2.87	3.84	2.81	2.80	2.78	2.77	2.75	2.75	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67	2.65	2.64	2.63	2.59	2.57	2.56	2.54	2.52	2.51	2.50	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50	2.49	2.48	2.46	2.42	2.40	2.38	2.36	2.35	2.33	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38	2.36	2.35	2.34	2.30	2.27	2.25	2.23	2.22	2.20	2.19	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28	2.27	2.26	2.24	2.20	2.17	2.16	2.13	2.12	2.10	2.09	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21	2.19	2.18	2.17	2.12	2.10	2.08	2.05	2.04	2.02	2.01	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15	2.13	2.12	2.10	2.06	2.03	2.01	1.99	1.97	1.95	1.94	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.12	2.10	2.08	2.07	2.05	2.01	1.98	1.96	1.93	1.92	1.89	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.07	2.05	2.04	2.02	2.01	1.96	1.93	1.91	1.89	1.87	1.85	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02	2.00	1.99	1.97	1.92	1.89	1.87	1.85	1.83	1.80	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	2.01	1.99	1.97	1.95	1.94	1.89	1.86	1.84	1.81	1.79	1.77	1.76	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.98	1.96	1.94	1.93	1.91	1.86	1.83	1.81	1.78	1.76	1.74	1.73	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.95	1.93	1.92	1.90	1.89	1.84	1.80	1.78	1.75	1.74	1.71	1.70	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.93	1.91	1.89	1.88	1.86	1.81	1.78	1.76	1.73	1.71	1.69	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.91	1.89	1.87	1.86	1.84	1.79	1.76	1.74	1.71	1.69	1.66	1.65	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.90	1.87	1.86	1.84	1.83	1.78	1.74	1.72	1.69	1.67	1.64	1.63	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86	1.84	1.83	1.81	1.76	1.73	1.70	1.67	1.65	1.63	1.61	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.87	1.84	1.83	1.81	1.80	1.74	1.71	1.69	1.66	1.64	1.61	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83	1.81	1.80	1.78	1.73	1.70	1.67	1.64	1.62	1.59	1.58	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.84	1.82	1.80	1.79	1.77	1.72	1.68	1.66	1.63	1.61	1.58	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.83	1.81	1.79	1.77	1.76	1.71	1.67	1.65	1.61	1.59	1.57	1.55	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.82	1.80	1.78	1.76	1.75	1.70	1.66	1.64	1.60	1.58	1.55	1.54	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79	1.77	1.75	1.74	1.69	1.65	1.63	1.59	1.57	1.54	1.53	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.80	1.78	1.76	1.75	1.73	1.68	1.64	1.62	1.58	1.56	1.53	1.52	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77	1.75	1.74	1.72	1.67	1.63	1.61	1.57	1.55	1.52	1.51	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.74	1.71	1.70	1.68	1.66	1.61	1.57	1.54	1.51	1.48	1.45	1.43	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66	1.64	1.62	1.60	1.54	1.50	1.48	1.44	1.41	1.38	1.36	1.29
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66	1.64	1.61	1.59	1.57	1.56	1.49	1.45	1.42	1.38	1.35	1.32	1.29	1.22
∞	2.71	2.30	2.08	1.95	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55	1.52	1.51	1.49	1.42	1.38	1.34	1.30	1.26	1.22	1.19	1.03

Tabla A5 Puntos porcentuales del estadístico rango estudentizado (5%).

$$q_{.05}\,(p, f)$$

	p														
f	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	18.10	26.70	32.8	37.20	40.50	43.10	45.40	47.30	49.10	50.60	51.90	53.20	54.30	55.40	56.30
2	6.09	8.28	9.80	10.89	11.73	12.43	13.03	13.54	13.99	14.39	14.75	15.08	15.38	15.65	15.90
3	4.50	5.88	6.83	7.51	8.04	8.47	8.85	9.18	9.46	9.72	9.95	10.16	10.35	10.52	10.60
4	3.93	5.00	5.76	6.31	6.73	7.06	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.67	8.80
5	3.61	4.54	5.18	5.64	5.99	6.28	6.52	6.74	6.93	7.10	7.25	7.39	7.52	7.64	7.75
6	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32	6.49	6.65	6.79	6.92	7.04	7.14	7.24
7	3.34	4.16	4.68	5.06	5.35	5.59	5.80	5.99	6.15	6.29	6.42	6.54	6.65	6.75	6.84
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57
9	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.60	5.74	5.87	5.98	6.09	6.19	6.28	6.36
10	3.15	3.88	4.33	4.66	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.12	6.20
11	3.11	3.82	4.26	4.58	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40	5.51	5.61	5.71	5.80	5.88	5.95
13	3.06	3.73	4.15	4.46	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.56	5.64	5.72	5.79
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72
16	3.00	3.65	4.05	4.34	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66
17	2.98	3.62	4.02	4.31	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.55	5.61
18	2.97	3.61	4.00	4.28	4.49	4.67	4.83	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57
19	2.96	3.59	3.98	4.26	4.47	4.64	4.79	4.92	5.04	5.14	5.23	5.32	5.39	5.46	5.53
20	2.95	3.58	3.96	4.24	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.50
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38
30	2.89	3.48	3.84	4.11	4.30	4.46	4.60	4.72	4.83	4.92	5.00	5.08	5.15	5.21	5.27
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74	4.82	4.90	4.98	5.05	5.11	5.17
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06
120	2.80	3.36	3.69	3.92	4.10	4.24	7.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95
∞	2.77	3.32	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.84

Tabla A6 Valores críticos para prueba de Duncan (1%).

$$r_{.01}\,(p, f)$$

	p											
f	2	3	4	5	6	7	8	9	10	20	50	100
1	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00
2	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00
3	8.26	8.50	8.60	8.70	8.80	8.90	8.90	9.00	9.00	9.30	9.30	9.30
4	6.51	6.80	6.90	7.00	7.10	7.10	7.20	7.30	7.30	7.50	7.50	7.50
5	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.50	6.80	6.80	6.80
6	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.00	6.30	6.30	6.30
7	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.80	6.00	6.00	6.00
8	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.50	5.80	5.80	5.80
9	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.40	5.70	5.70	5.70
10	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.55	5.55	5.55
11	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.39	5.39	5.39
12	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.26	5.26	5.26
13	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.15	5.15	5.15
14	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	5.07	5.07	5.07
15	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	5.00	5.00	5.00
16	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.94	4.94	4.94
17	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.73	4.75	4.89	4.89	4.89
18	4.07	4.27	4.39	4.46	4.53	4.59	4.64	4.68	4.71	4.85	4.85	4.85
19	4.05	4.24	4.36	4.43	4.50	4.56	4.61	4.64	4.67	4.82	4.82	4.82
20	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.79	4.79	4.79
30	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.65	4.71	4.71
40	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.59	4.69	4.69
60	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.53	4.66	4.66
100	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.48	4.64	4.65
∞	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.41	4.60	4.68

f = grados de libertad.

Tabla A6 (*Continuación*) Prueba de Duncan 5%.

$$r_{.05}(p, f)$$

	p											
f	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**20**	**50**	**100**
1	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.25	3.52	3.52	3.52	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.48	3.48	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.48	3.48	3.48
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.48	3.48	3.48
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.47	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.47	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.47	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.47	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.47	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.47	3.47	3.47
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.47	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.47	3.47	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.47	3.47	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.47	3.47	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.47	3.48	3.48
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.47	3.47	3.53
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.47	3.61	3.67

f = grados de libertad.

Tabla A7 Valores críticos para la prueba de Dunnett. (Hipótesis bilaterales.)

	$a - 1$ = Número de medias de tratamiento (sin contar el control)								
f	1	2	3	4	5	6	7	8	9
5	2.57	3.03	3.29	3.48	3.62	3.73	3.82	3.90	3.97
6	2.45	2.86	3.10	3.26	3.39	3.49	3.57	3.64	3.71
7	2.36	2.75	2.97	3.12	3.24	3.33	3.41	3.47	3.53
8	2.31	2.67	2.88	3.02	3.13	3.22	3.29	3.35	3.41
9	2.26	2.61	2.81	2.95	3.05	3.14	3.20	3.26	3.32
10	2.23	2.57	2.76	2.89	2.99	3.07	3.14	3.19	3.24
11	2.20	2.53	2.72	2.84	2.94	3.02	3.08	3.14	3.19
12	2.18	2.50	2.68	2.81	2.90	2.98	3.04	3.09	3.14
13	2.16	2.48	2.65	2.78	2.87	2.94	3.00	3.06	3.10
14	2.14	2.46	2.63	2.75	2.84	2.91	2.97	3.02	3.07
15	2.13	2.44	2.61	2.73	2.82	2.89	2.95	3.00	3.04
16	2.12	2.42	2.59	2.71	2.80	2.87	2.92	2.97	3.02
17	2.11	2.41	2.58	2.69	2.78	2.85	2.90	2.95	3.00
18	2.10	2.40	2.56	2.68	2.76	2.83	2.89	2.94	2.98
19	2.09	2.39	2.55	2.66	2.75	2.81	2.87	2.92	2.96
20	2.09	2.38	2.54	2.65	2.73	2.80	2.86	2.90	2.95
24	2.06	2.35	2.51	2.61	2.70	2.76	2.81	2.86	2.90
30	2.04	2.32	2.47	2.58	2.66	2.72	2.77	2.82	2.86
40	2.02	2.29	2.44	2.54	2.62	2.68	2.73	2.77	2.81
60	2.00	2.27	2.41	2.51	2.58	2.64	2.69	2.73	2.77
120	1.98	2.24	2.38	2.47	2.55	2.60	2.65	2.69	2.73
∞	1.96	2.21	2.35	2.44	2.51	2.57	2.61	2.65	2.69

f = grados de libertad.

Tabla A7 (*Continuación*) Valores críticos para la prueba de Dunnett (hipótesis unilaterales).

$a - 1$ = Número de medias de tratamiento (sin contar el control)									
f	1	2	3	4	5	6	7	8	9
5	2.02	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30
6	1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
7	1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.95	3.01
8	1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
9	1.83	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86
10	1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
11	1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
12	1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
13	1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
14	1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
15	1.75	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
16	1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
17	1.74	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
18	1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
19	1.73	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
20	1.72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
24	1.71	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57
30	1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
40	1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
60	1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
120	1.66	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
∞	1.64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42

Tabla A8 Límites para prueba de Durbin-Watson.

Nivel de significancia $\alpha = .05$, $p - 1$ es el número de variables o términos en el modelo, n el número de datos										
	$p-1=1$		$p-1=2$		$p-1=3$		$p-1=4$		$p-1=5$	
n	d_L	d_v	d_L	d_v	d_L	d_v	d_L	d_v	d_L	d_v
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.45	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.72	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Apéndice B

Uso de sistemas computacionales

Sumario

- Etapas al planear y analizar un experimento en un paquete estadístico
- Sistema *Minitab*
- Sistema *JMP*
- Diseño de experimentos usando *SPSS*

Como se ha comentado, en estadística es fundamental el uso de sistemas computacionales especializados para mejorar y facilitar la planeación y el análisis de los estudios experimentales. Por ello, al final de cada capítulo se agregó una sección en la cual se explica cómo utilizar algunos sistemas computacionales. En este sentido, se hizo énfasis en *Statgraphics* y en *Minitab*, aunque en algunos capítulos también se habla de *Excel* y *Design Expert*.

En este apéndice se presenta un panorama introductorio de *Statgraphics* y *Minitab*, además se dan algunas indicaciones de cómo utilizar *JMP* y *SPSS*, otros dos sistemas computacionales que tienen una presencia importante.

Etapas al planear y analizar un experimento en un paquete estadístico

A fin de mantener un orden al momento de planear y analizar un experimento con cualquier paquete estadístico, conviene clasificar las diferentes actividades en cuatro etapas, como se muestra en la tabla B1. Para cada etapa se plantean algunas preguntas relevantes, las cuales determinan las salidas del paquete que son relevantes en la etapa. La idea es evitar la frecuente situación en que el usuario no tiene claridad en lo que desea conseguir del paquete que está usando, lo cual ocasiona que observe salidas que no son relevantes en la etapa que está analizando.

Sistema *Statgraphics*

Cuando inicia este sistema normalmente aparece la pantalla de *StatWizard*, que ayuda a realizar la tarea deseada, ya sea introducir nuevos datos o realizar algún análisis específico, como por ejemplo, diseñar un nuevo experimento. En la figura B1 se muestra la pantalla general del software en su versión anterior y en la versión *Centurion* (una versión con cambios en los menús).

Para introducir una nueva variable o nuevos datos, se selecciona una columna dando clic con el puntero donde está *Col_1*, después, al presionar el lado derecho del *mouse* aparecen varias opciones (ver recuadro inferior de la figura B1), entre ellas la de *Modify Column*. Al seleccionarla aparecerá una ventana donde se podrá dar el nombre de la variable y especificar de qué tipo es (*Numeric*, *Character*, etc.). Después de eso se da Enter y se procede a introducir los datos, o bien, a especificar las características de otra variable en una segunda columna.

Por ejemplo, en la figura B2 se muestra cómo se definió el archivo Envases.sf3 y la variable (columna) resistencia, en donde se observa parte de los datos de la resistencia de un envase de plástico en posición vertical. Primero se hará un análisis exploratorio de estos datos para ver cómo es la resistencia de este producto. Para ello se sigue la secuencia: *Describe* → *Numeric Data* → *One-Variable Analysis*, y se declara la variable a analizar (véase figura B2). Después de eso, en todos los procedimientos de *Statgraphics* aparecerá una pantalla como la de la figura B3, donde se podrá seleccionar una serie de análisis en forma de tabla (*Tables*) o en forma gráfica (*Graphs*). En la figura B3 se muestran los análisis contemplados dentro de *One-*

Tabla B1 Etapas del diseño de experimentos y sus salidas relevantes.

Etapa	Preguntas en la etapa	Salidas y opciones relevantes
1. Planeación y diseño	• ¿Cuántos factores? • ¿Cuántas respuestas? • ¿Cuántos niveles?, ¿cuáles niveles? • ¿Cuántas réplicas o repeticiones? • ¿Hay restricciones de algún tipo?, ¿en cuántos bloques? • ¿Cuánto cuesta cada corrida? • ¿Cuáles efectos se podrán estudiar? • ¿Cuál es el experimento más apropiado? • ¿Cómo se capturan los datos?	Sólo algunos aspectos de esta etapa se trabajan en el software, pero hay más opciones si el diseño es un factorial fraccionado: determinación del tamaño de muestra o número de réplicas, matriz de diseño aleatorizada, generadores, resolución, estructura de alias, repeticiones, puntos al centro, bloques y hoja de trabajo para capturar los datos del experimento.
2. Análisis	• ¿Cuáles efectos son significativos? • ¿Cuáles factores contribuyen a explicar la respuesta? • ¿Cuánto explica cada modelo? • ¿Cuál es el modelo más apropiado?	Análisis de varianza (ANOVA), significancia de los efectos, gráfico de efectos en papel normal, Pareto de efectos, agregar y excluir efectos, porcentaje de explicación de cada modelo y selección del mejor modelo.
3. Interpretación	• ¿Cuál es el mejor tratamiento? • ¿Cómo afecta el factor *X*? • ¿Cuál es el valor esperado de la respuesta en el mejor tratamiento? • ¿Cómo es la superficie de respuesta? • ¿Qué porcentaje de la variación observada explica el modelo? • ¿Cuál es el tratamiento óptimo?	Gráficas de medias, gráficas de efectos, tabla de predichos o estimados, gráfico de cubo, intervalos de confianza para la respuesta media y para la observación futura, gráficas de superficie y de contornos, porcentaje de explicación de la respuesta, optimización, determinación del mejor tratamiento o punto óptimo.
4. Diagnóstico	• ¿Se cumple los supuestos de normalidad, varianza constante e independencia? • ¿Hay observaciones atípicas? • ¿Se ajusta bien el modelo?	Gráficas de residuos: en papel de probabilidad normal, contra los predichos, contra los factores y ordenados en el tiempo. Prueba de homogeneidad de varianzas, observaciones influyentes y gráfico de observados contra predichos.

Variable Analysis. Por ejemplo, en la tabla B2 se muestra *Summary Statistics* para resistencia.

Ésta contiene medidas de tendencia central, variabilidad y forma de la distribución de los datos. Por lo general, en cada *Pane* hay opciones relativas al análisis correspondiente; por ejemplo, en el caso del análisis de los datos podría ser de interés calcular otros estadísticos (moda, mediana, percentiles, etc.), a los que se puede acceder desde el *Pane* de *Summary Statistics*, usando el botón derecho del puntero (ratón) y seleccionando *Pane Options*, donde es posible elegir los estadísticos que se desea calcular.

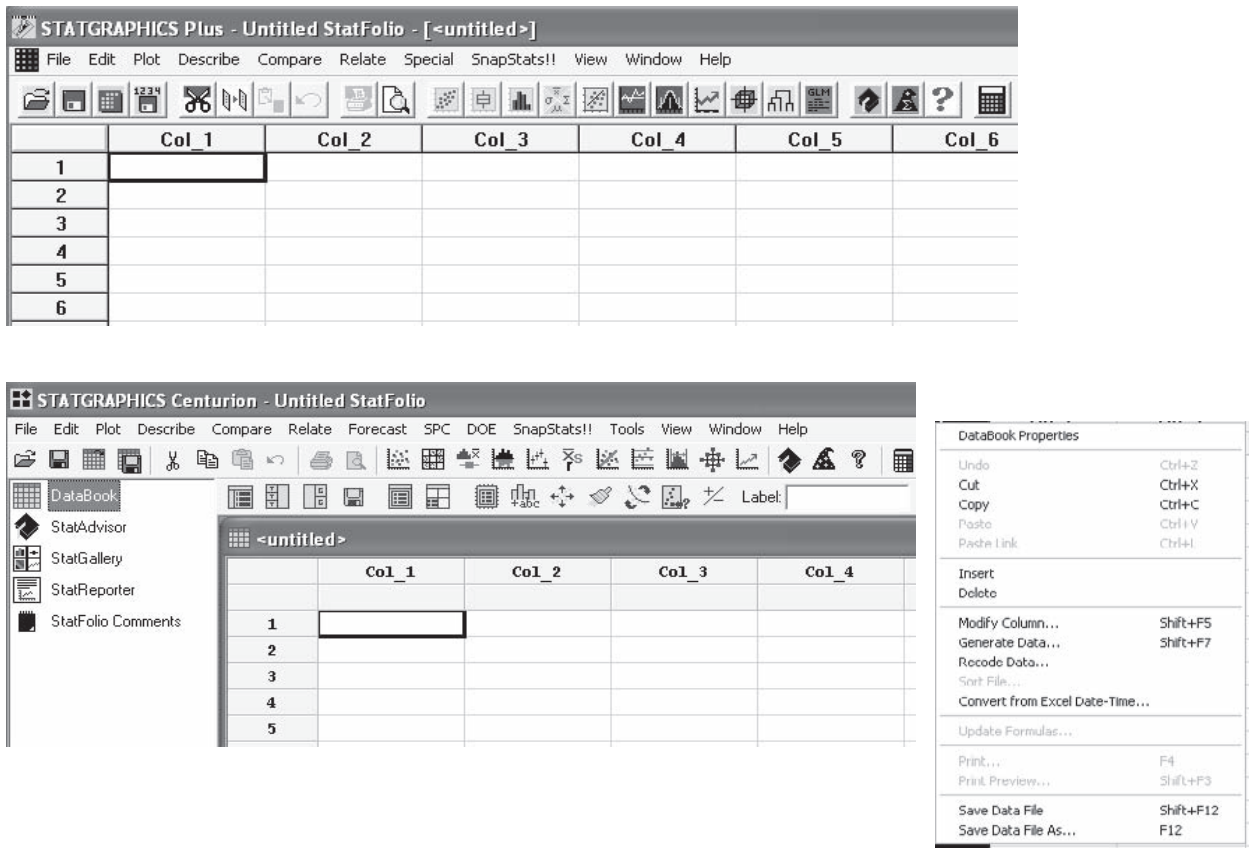


Figura B1 Pantallas iniciales para las dos versiones de *Statgraphics*.

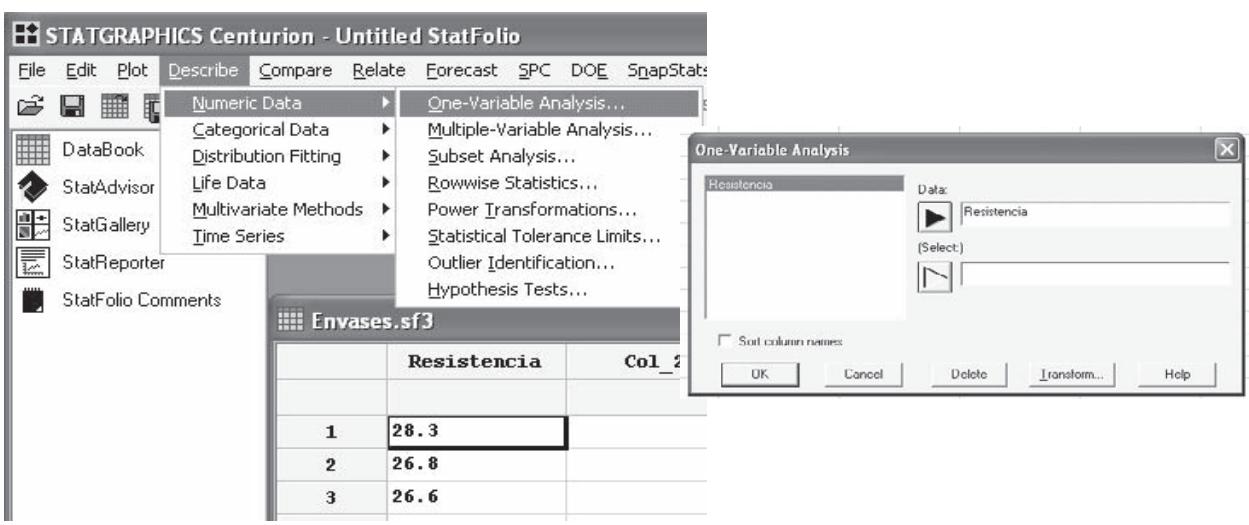


Figura B2 Secuencia para hacer un análisis descriptivo de un conjunto de datos en *Statgraphics*.

Tabla B2 Tabla de *Summary Statistics* para resistencia.

Count	56
Average	27.2464
Standard deviation	1.43044
Coeff. of variation	5.25002%
Minimum	23.7
Maximum	30.4
Range	6.7
Stnd. skewness	–0.467625
Stnd. kurtosis	–0.52401

Tipos de archivos

Los análisis hechos en *Statgraphics* se pueden grabar en alguno de los siguientes tipos de archivos:

- *Data File*. En este tipo de archivos sólo se guardan los datos que están en la tabla de datos.
- *StatFolio*. Aquí se guardan todas las gráficas y tablas que se generan en una sesión de trabajo.
- *StatReporter*. Al estar en cualquier *Pane*, con el botón derecho del puntero (ratón) se puede copiar en un archivo de este tipo, ya sea el *Pane* o todo el análisis.

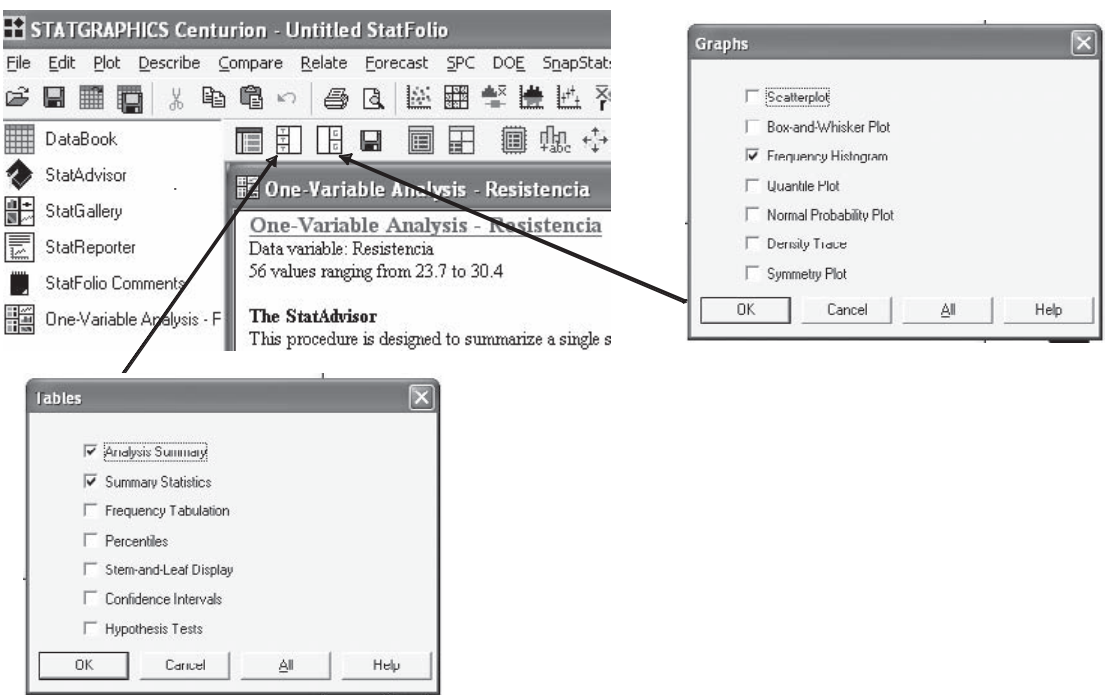


Figura B3 Opciones de tablas y gráficas en *One-Variable Analysis* en *Statgraphics*.

Este archivo tiene formato rtf, que puede abrirse con muchos procesadores de texto (por ejemplo con Microsoft Word). Así que esta opción se utiliza cuando se quiere trabajar con las tablas y gráficas de un análisis en otro procesador, ya sea para hacer un reporte o una exposición.

- *StatGallery*. Aquí se pueden copiar diferentes gráficas de uno o varios análisis con la finalidad de verlas juntas y arreglarlas en un formato adecuado (2×2, 3×3, etc.). Para copiar una gráfica en *StatGallery* es necesario posicionarse en la gráfica y presionar el botón derecho del puntero (ratón), con lo que aparecerán varias opciones, entre ellas la de *Copy Pane to StatGallery*. Después, simplemente se pega la gráfica en la parte deseada de *StatGallery*.

Algo de particular utilidad en este software es que debajo de cada tabla de un análisis siempre aparece una ayuda para interpretar la salida (*The StatAdvisor*). Se recomienda al lector siempre procurar la lectura de este análisis.

Versión en español

Cabe señalar que cuando se escribió este apéndice, ya existían versiones de prueba de este software completamente en español, lo que sin lugar a dudas facilitará su mejor aprovechamiento por parte de la comunidad de habla hispana de todo el mundo.

Le recomendamos al lector que antes de utilizar el software para realizar alguno de los procedimiento descritos en este libro, lea la sección final (antes de los ejercicios) de cada capítulo, donde se dan indicaciones que le serán de utilidad.

Sistema *Minitab*

Minitab es un paquete estadístico que permite aplicar una gran variedad de técnicas estadísticas para el análisis de datos de diversa índole, como se aprecia en la opción *Stat* del menú principal del software que se muestra en la figura B4. Al iniciar *Minitab* aparecen las dos ventanas de la figura B4. En la ventana inferior, tipo hoja de cálculo, se capturan los datos, y en la ventana superior el paquete mostrará todas las salidas numéricas de los análisis que se realicen. Para cada salida gráfica el paquete abre una ventana adicional. En la figura B4 se muestra el menú de la opción ANOVA, donde se analizan los diseños comparativos con más de dos tratamientos o poblaciones a comparar. Los análisis para una o dos poblaciones se encuentran en la opción *Basic Statistics*.

Del esquema dado en la tabla B1, la etapa de planeación y diseño para los diseños comparativos, en este paquete se reduce básicamente a la determinación del tamaño de muestra en la opción *Power and simple size* y a la captura de los datos en la hoja de trabajo. En cuanto a las etapas siguientes, el análisis de varianza del diseño completamente al azar o del diseño en bloques completos al azar se realiza en ANOVA en las opciones *One Way* y *Two Way*, respectivamente. Las gráficas de interpretación de los resultados se encuentran en la última parte del menú de ANOVA (véase figura B4). Ahí mismo, en la opción *Graphs*, se pueden pedir las gráficas de residuos para el diagnóstico del modelo.

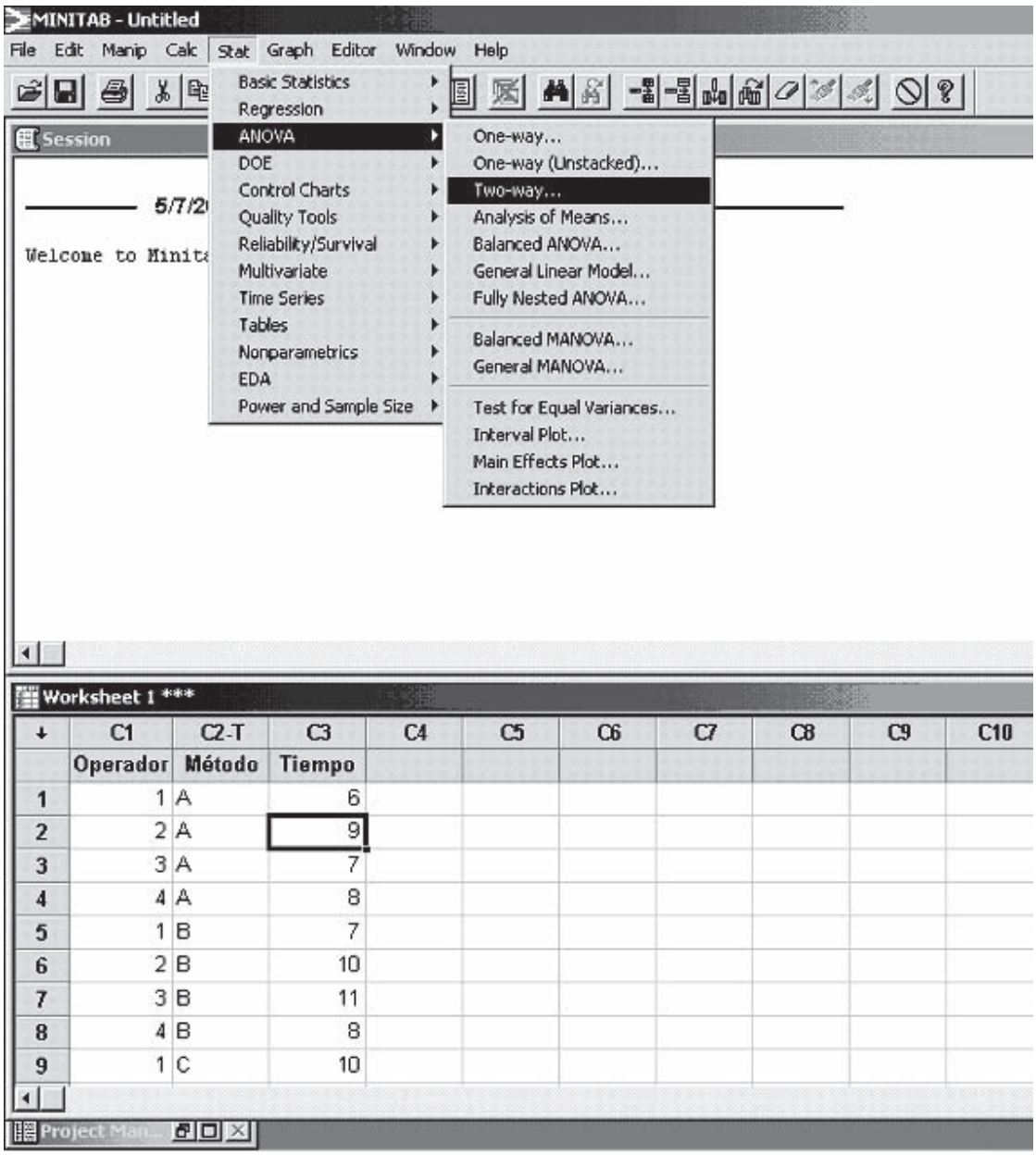


Figura B4 Ambiente de Minitab con los menús de *Stat* → ANOVA.

Diseño factorial

En la opción de *Doe* → *Factorial* → *Create Factorial Design*, se despliega la ventana dada en la parte izquierda de la figura B5, donde las características del diseño factorial deseado se definen de la siguiente manera: primero se especifica el número de factores, luego en *Display Available Designs* se despliegan los diseños disponibles clasificados de manera práctica de acuerdo al número de factores y al número de corridas y su resolución. En *Designs* se selecciona el diseño base deseado, al cual se le agregan ahí mismo los bloques, las réplicas y los puntos al centro por bloque.

Una vez seleccionado un diseño, se activan los botones tenues de la ventana izquierda de la figura B5, los cuales permiten completar la creación del diseño final deseado: en *Factors* se definen los nombres y niveles de los factores a estudiar; mientras que en *Options* se elige la fracción deseada, si ése fuera el caso, y se decide ahí mismo la aleatorización de las corridas y la generación de la matriz de diseño final en la hoja de datos en el orden (aleatorio) especificado. Por último, se corre el diseño y se capturan los datos en la hoja de trabajo.

Etapa de interpretación

La salida de coeficientes y la tabla de ANOVA en la opción *Results* de la ventana derecha de la figura B5, también es relevante en esta etapa, ya que ahí se muestran los efectos estimados. Las medias estimadas en cada nivel de cada efecto se pueden pedir en la última opción dentro de *Results*. En cuanto a las salidas gráficas, éstas se encuentran en una opción especial con la secuencia: *Doe* → *Factorial* → *Factorial Plots*. En la ventana que se genera es posible elegir gráficas de efectos principales, de efectos de interacción y gráficos de cubo, en términos de medias ajustadas u observadas. Los gráficos de contornos y de superficie también tienen su propia secuencia especial que es: *Doe* → *Factorial* → *Contour/Surface* (*Wireframe*) → *Plots*.

Sistema *JMP*

En la figura B6 se muestra el menú principal de este software con los datos de resistencia de envases. Por medio de la pestaña de *Window* se accede al *JMP Starter*, un software que ayuda a localizar diversos análisis. En la segunda parte de la figura B6 se muestra el *JMP Starter* con la activación de la pestaña de *Doe* (diseño de experimentos). Como se aprecia, aparecen los diseños vistos en este libro. Por ejemplo, si elige *Full Factorial Design* (ver capítulo 5) aparece una pantalla donde será necesario especificar el número y características de las variables de respuesta y cuántos

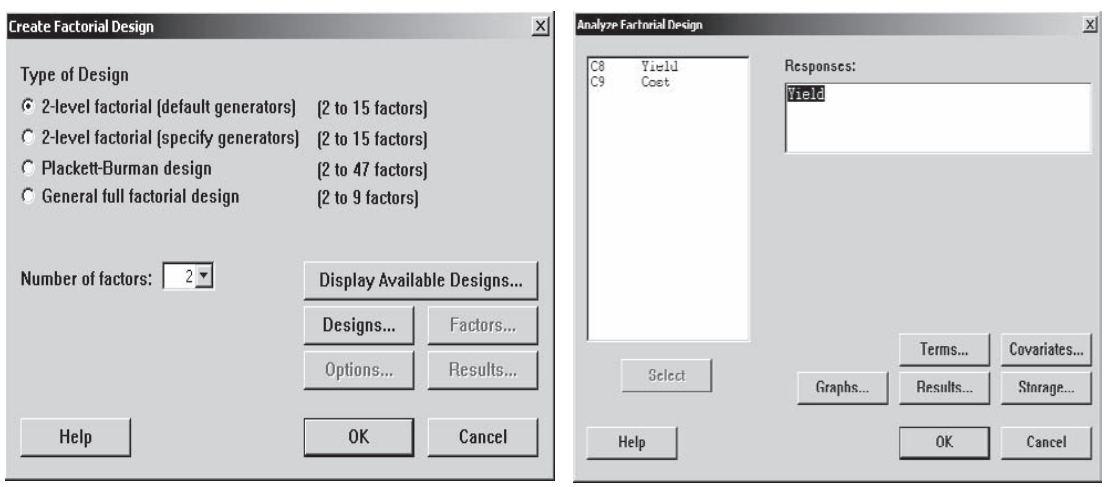


Figura B5 Creación y análisis del diseño factorial en *Minitab*.

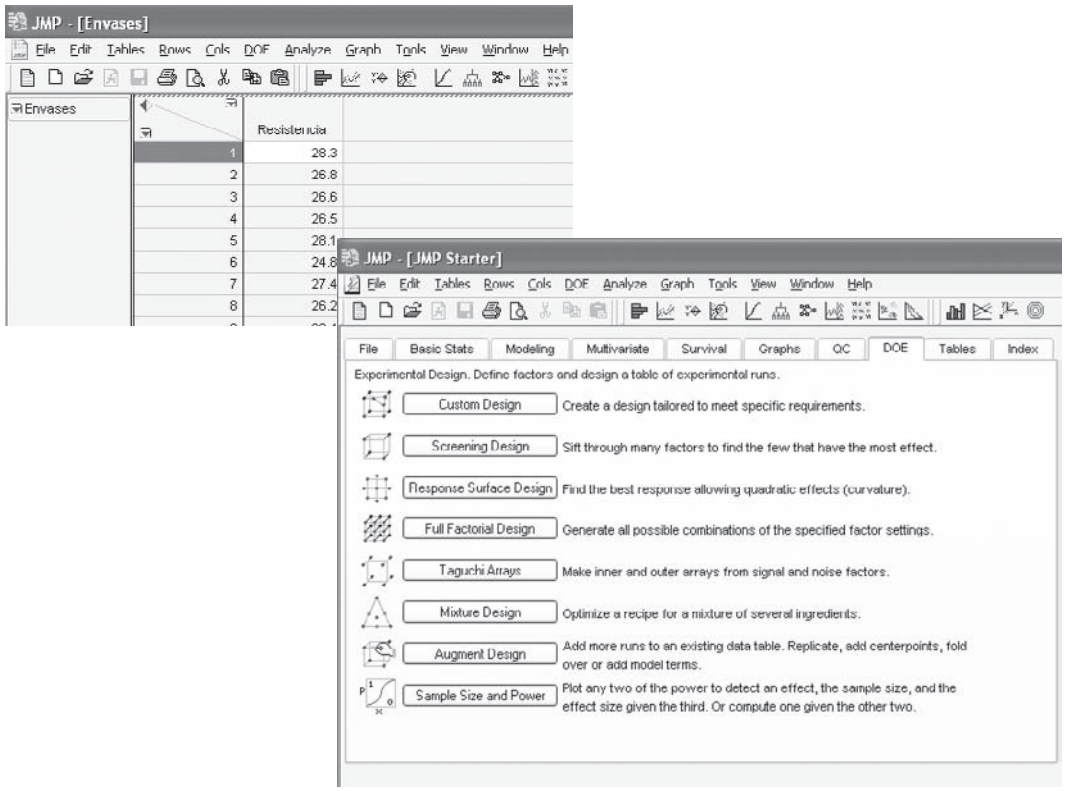

JMP - [Envases]
File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help
Envases
Resistencia
1 28.3
2 26.8
3 26.6
4 26.5
5 28.1
6 24.8
7 27.4
8 26.2
JMP - [JMP Starter]
File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help
File Basic Stats Modeling Multivariate Survival Graphs QC DOE Tables Index
Experimental Design. Define factors and design a table of experimental runs.
Custom Design
Create a design tailored to meet specific requirements.
Screening Design
Sift through many factors to find the few that have the most effect.
Response Surface Design
Find the best response allowing quadratic effects (curvature).
Full Factorial Design
Generate all possible combinations of the specified factor settings.
Taguchi Arrays
Make inner and outer arrays from signal and noise factors.
Mixture Design
Optimize a recipe for a mixture of several ingredients.
Augment Design
Add more runs to an existing data table. Replicate, add centerpoints, fold over or add model terms.
Sample Size and Power
Plot any two of the power to detect an effect, the sample size, and the effect size given the third. Or compute one given the other two.

Figura B6 Menú principal de *JMP* y el *JMP Starter* con la opción de *Doe* (diseño de experimentos).

factores de tipo continuo y categórico se tienen en el estudio experimental, y cuántos niveles en cada caso. Después se define el número de réplicas y si se aleatoriza (*randomize*) el orden de las corridas experimentales. Con ello, enseguida se genera la matriz de diseño correspondiente. Después de que se registran los valores de las variables de respuesta, para hacer el análisis se selecciona la pestaña de *Analyze*, y luego la opción *Fit Model*, después aparecerá una pantalla con el modelo a ajustar (que es apropiado para el diseño de experimentos corrido). Entonces, se corre el modelo (pestaña *Run Model*) para que aparezcan los análisis apropiados para cada variable de respuesta. En los análisis que aparece un *triángulo invertido de color rojo*, es una indicación de que existen opciones adicionales para esa gráfica o tabla.

Para hacer un reporte de todo el análisis realizado en un archivo con formato *rtf*, que puede ser leído con procesadores de palabras, como *Microsoft Word*, una vez realizado el análisis se selecciona la pestaña *Edit* en el menú principal y se elige *Journal*, después con la pestaña *File* y *Save as* se graba este archivo, teniendo la precaución de elegir el formato deseado.

Diseño de experimentos usando *SPSS*

El sistema *SPSS* no ofrece un menú especializado para generar o analizar diseños de experimentos, de manera que varias actividades de planeación, como es la genera-

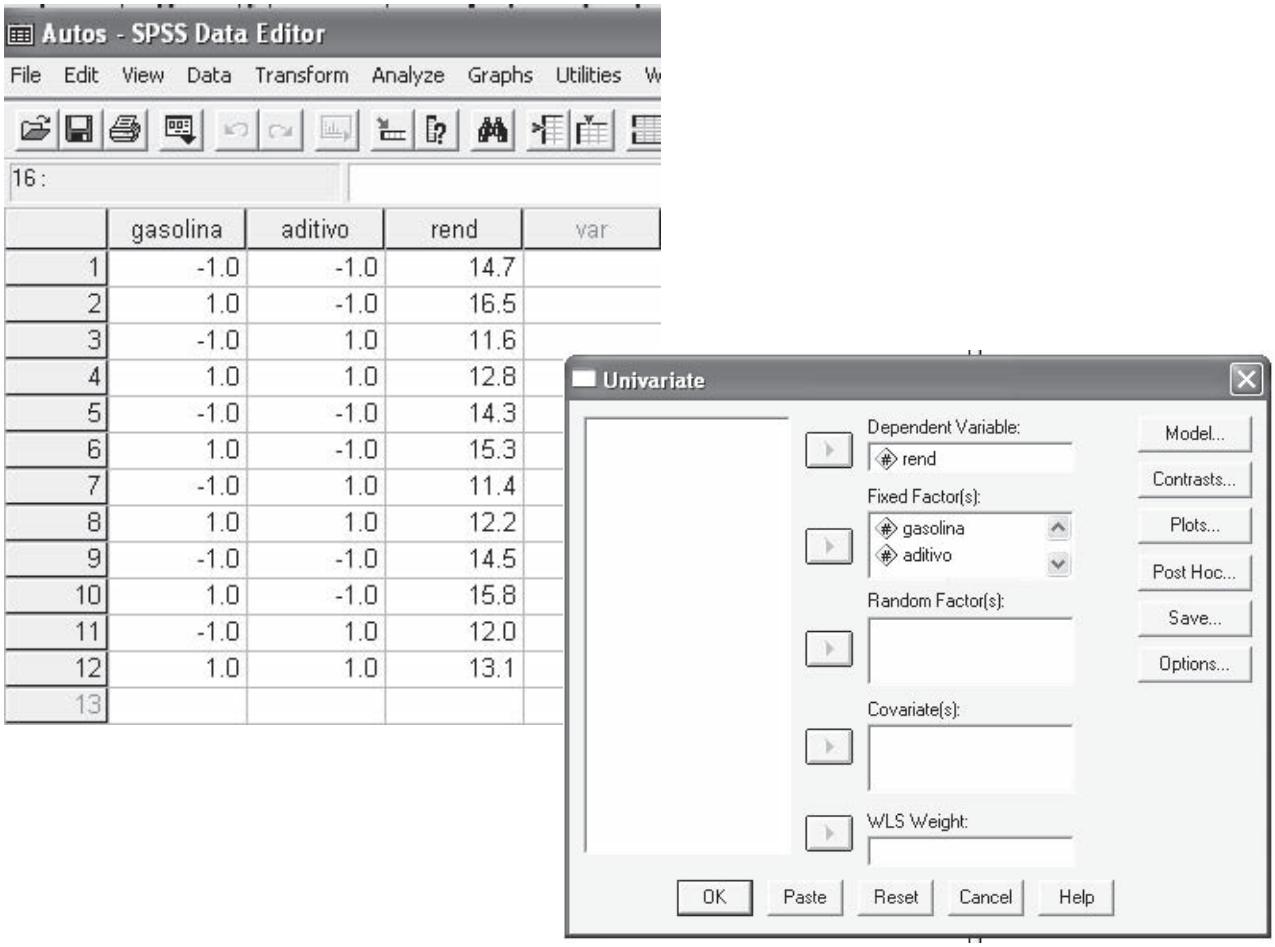


Figura B7 Pantallas de inicio de *SPSS*, introducción de datos y modelos lineales generalizados univariados en *SPSS*.

ción de un experimento factorial, se hacen de manera manual o con algún otro software. Sin embargo, este paquete cuenta con las opciones estadísticas de uso genérico suficientes para analizar todo tipo de experimentos.

La pantalla de entrada de *SPSS* se presenta en lado izquierdo de la figura B7. En la hoja de trabajo que se muestra en la figura, llamada *Data View*, se incluyen tres variables que corresponden a un experimento factorial 2^2 con tres réplicas, donde se estudia el rendimiento en km/litro en función del tipo de gasolina (magna, premium) y la marca de aditivo. Estos datos se capturan de forma directa, pero previamente se definen las variables, los niveles y los nombres de éstos en la hoja *Variable View*, que es una segunda hoja tipo *Excel* cuya ceja para ponerla en primer plano se aprecia en la parte inferior de la hoja de datos (*Data View*) de la figura B7. En la primera columna de la hoja *Variable View* se listan todas las variables involucradas en el experimento.

Una vez que se introducen los datos del experimento se procede con el análisis. Si el experimento es un diseño completamente al azar su análisis se realiza en la opción: *Analyze* → *Compare means* → *One-way ANOVA*. Pero si es un diseño de bloques o factorial se analiza en: *Analyze* → *General linear model* → *Univariate* (ver lado

derecho de la figura B7). Ahí, la opción *Model* permite especificar el modelo deseado, mientras que la opción *Contrasts* permite probar la significancia de combinaciones lineales de los niveles para cada factor. Las pruebas de comparaciones múltiples (LSD, Tukey, etc.) son parte de la opción *Post Hoc*. En la opción *Save* se pueden pedir diferentes tipos de residuos, los predichos, medidas de influencia y mandar a un archivo nuevo los efectos estimados. En *Options* es posible pedir estadísticos adicionales de varios tipos, como por ejemplo, parámetros estimados, prueba de falta de ajuste, prueba de homogeneidad de varianzas y estadísticas descriptivas para cada tratamiento o combinación de niveles. En la opción de *Plots* se pueden graficar los efectos principales y de interacción definiendo tres gráficos separados, para ello se oprime tres veces el botón *Add* que está dentro de *Plots*; pero para la interacción habría que declarar el segundo factor en el campo llamado *Separate lines*.

Referencias y bibliografía

Anderson, M. J. y Whitcomb, P. J. (2002). Mixture DOE uncovers formulations quicker. *Rubber and Plastics News*, 21 de octubre, pp. 16-18.

Arroyo-Gutiérrez, V., Becerril Mondragón, B. E., Cantú-Ortega, D. M y B. E. Hernández-Hernández (2003). Esponjosidad en Hotcakes. Proyecto final de diseño de experimentos, IIMAS, UNAM, México.

Beckman, R. J., Nachtsheim, C. J. y Cook, R. D. (1987). Diagnostic for mixed-model analysis of variance. *Technometrics* 29, pp. 413-426.

Bingham, D. R., Schoen, E. D. y Sitter, R. R. (2004). Designing fractional factorial split-plot experiments with few whole-plot factors. *Applied Statistics* 53, Parte 2, pp. 325-339.

Bisgaard, S. y Fuller, H. T. (1995). Sample Size Estimates for 2^(k-p) Designs with Binary Responses. *Journal of Quality Technology*, vol. 27, pp. 344-354.

Bisgaard, S. y Kulahci, M. (2001). Robust product design: saving trials with split-plot confounding. *Quality Engineering* 13, pp. 525-530.

Bowles, M. L. y Montgomery, D. C. (1998). How to Formulate the Ultimate Margarita: A Tutorial on Experiments with Mixtures. *Quality Engenering* 10, 2, pp. 239-253.

Box, G. (1988) Signal-to-noise ratios, performance criteria and transformation. Technometrics, **30**(1), pp. 1-17.

Box, G. E. P. (1988). "Signal to noise ratios, performance criteria, and transformations". *Technometrics*, 30, pp. 1-17.

Box, G. E. P. (1993). "Quality improvement. The new industrial revolution", *International Statistical Review*, 61, 3-19.

Box, G. E. P. y Jones, S. (2000). Split plots for robust product an process experimentation. *Quality Engineering* 13, pp. 127-134.

Box, G. E. P. y K. B. Wilson (1951). "On the experimental attainment of optimum conditions (with discussion)". *Journal of the Royal Statistical Society*, B, 13, 1-45.

Box, G. E. P. y N. R. Draper (1969). *Evolutionary operation*. Wiley, Nueva York.

Box, G. E. P. y N. R. Draper (1987). *Empirical model-building and response surfaces*. Wiley, Nueva York.

Box, G. E. P. y P. Y. T. Liu (1999). "Statistics as a catalyst to learning by scientific method". *Journal of Quality Technology*, vol. 31, núm. 1 pp. 1-29.

Box, G. E. P., Hunter W. G. y J. S. Hunter (1978). *Statistics for experimenters*. Wiley, Nueva York.

Byrne, D. M. y Shin Taguchi (1987). "The Taguchi approach to parameter design". *Quality Progress*, diciembre.

Charbonneau, H. C. y Webster, G. L. (1983). *Control de calidad*. Editorial Interamericana, México.

Chen, J., Sun, D. X. y C. F. J. Wu (1993). "A catalogue of two-level and three-level fractional designs with small runs". *International Statistical Review*, 61, 131-145.

Cochran, W. G. y G. M. Cox (1983). *Diseños experimentales*. Trillas, México.

Conover, W. J. (1980). *Practical nonparametric statistics*, 2a. ed., John Wiley, Nueva York.

Contreras, V. y Ochoa, A. (2003). Estacas cortas y gruesas: una opción agronómica para la siembra vegetativa de *Gliricidia sepium*. *Zootecnia Tropical* 21, 413-423.

Cornell, J. A. (1990). *Experiments with mixtures*. Wiley, Nueva York.
Cornell, J. A. (2002). *Experiments with Mixtures*. 3a. ed., John Wiley, Nueva York.
Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. Technometrics **1**, 311-341.
De la Vara, R. (1995). "Un experimento exitoso en la fabricación de circuitos integrados". *Revista de Estadística*, VII, 53-63.
De la Vara, R. y J. Domínguez (1988). "Metodología de superficie de multirrespuesta". *Comunicación Técnica*, I-90-08, CIMAT.
De la Vara, R. y J. Domínguez (2002). Métodos de superficie de multirrespuesta: un estudio comparativo. *Revista Matemática: Teoría y Aplicaciones* 9, 47-65.
Dean, A. y Voss, D. (1999). *Design and Analysis of Experiments*. Springer-Verlag, Nueva York.
Del Castillo, E., Montgomery, D. C. y McCarville, D. R. (1996). Modified desirability functions for multiple response optimization. *Journal of Quality Technology* 28, pp. 337-345.
Deming, W. E. (1989). *Calidad, productividad y competitividad*. Madrid, Díaz de Santos.
Derringer, G. C. (1990). "A balancing act: Optimizing a product's properties", *Quality Progress*, junio de 1994.
Derringer, G. C. y R. Suich (1980). "Simultaneous optimization of several responses". *Journal of Quality Technology*, 12, 214-219.
Design Expert. *Software for design of experiments*. Version 6.0.2. Stat.Ease, Inc.
Draper, N. R. y Lin, D. K. J. (1990). Small Response-Surface Designs. *Technometrics*, vol. 32, pp. 187-194.
Espinoza Escalante, F. M., Pelayo Ortiz, C., Bories A., Gutiérrez Pulido, H., González Álvarez, V. y Alcaraz González, V. (2006). Análisis estadístico de diseño de mezclas y optimización múltiple aplicado al proceso de digestión anaerobia de vinazas tequileras. Bol. Nakari 2006, 17(3), pp. 75-80.
Fowlkes, W. Y. y Creveling, C. M. (1995). *Engineering methods for robust product design*. Addison Wesley, USA.
Frisbee, S. E. y McGinity, J. W. (1994). Influence of nonionic surfactants on the physical and chemical properties of a biodegradable pseudolatex. *European Journal of Pharmaceutics and Biopharmaceutics* 40, diciembre, núm. 6.
García. T. y Espinoza, J. (2006). Evaluación de los factores que inciden en el comportamiento reológico de una suspensión de esmalte para baldosas mediante el uso de experimentos de mezcla con espacio de mezcla restringido. *Memorias del XVI Simposio de Estadística*, Universidad Nacional de Colombia.
Grize, Y. L. (1995). "A review of robust process design approaches". *Journal of Chemometrics*, 9, pp. 239-262.
Gutiérrez-Pulido, H. (1997). *Calidad total y productividad*. McGraw-Hill, México.
Gutiérrez-Pulido, H. y De la Vara Salazar, R. (2003). *Control estadístico de calidad y Seis Sigma*. En prensa, McGraw-Hill, México.
Hao, X.-C., Xiao-Bin Yu, X-B. y Yan, Z.-L. (2006). Production of cellulase by trichoderma reesei WX-112. *Food Technology Biotechnology*, 44 (1), pp. 89-94.
Harrington, E. (1965). The desirability function. *Industrial Quality Control* 21, pp. 494-498.
Hicks, C. R. (1982). *Fundamental concepts in the design of experiments*. 3a. ed., Holt, Rinehart and Winston, Nueva York.
Hinkelmann, K. y O. Kempthorne (1994). *Design and analysis of experiments. Volume I: Introduction to experimental design*. Wiley, Nueva York.
Hunter, J. S. (1989). "Let's all beware the latin square", *Quality Engineering*, 1, 453-465.
Kachar, R. N. (1985). "Off-line quality control, parameter design and the Taguchi method". *Journal of Quality Technology*, 17, pp. 176-209.

Khuri, A. I. y Cornell, J. A. (1987). *Response surfaces: design and analyses*. Nueva York: Marcel Dekker.

Kowalski, S. M. y Potcner, K. J. (2003). How to recognize a split-plot experiment. *Quality Progress* 36, núm. 11, pp. 60-66.

Langford, H. G., Davis, B. R., Blaufox, D., Oberman, A., Wassertheil-Smoller, S., Hawkins, M. y N. Zimbaldi (1991). *Hypetension* 17, pp. 210-217.

Lawson, J., Madrigal, J. I. y J. Erjabvec (1992). *Estrategias experimentales para el mejoramiento de la calidad en la industria*. Grupo Editorial Iberoamérica, México.

Lobato Calleros, C., Vernon C., E. J., Guerrero L., J., Soriano M., J. y Escalona B., H. (1997). Efecto del tipo de grasa sobre las características texturales instrumentales y sensoriales de análogos de queso. *Dintel*, año 3, núm. 4, pp. 57-67. Universidad Autónoma de Chapingo, México.

Lochner, R. H. y J. E. Matar (1990). *Designing for quality*, ASQC Quality Press, Milwaukee, Wisconsin.

Logothetis, N. y Wynn, H. P. (1994). *Quality through design*. Clarendon Press, Oxford.

McLean, R. A. y Anderson, V. L. (1966). Extreme vertice design of mixture experiments, *Technometrics*, 8, pp. 447-454.

McNeese, W. H. y Klein, R. A. (1991). *Statistical methods for the process industries*. ASQC Quality Press, Milwaukee, Wisconsin.

Minitab Statistical Software. Minitab Inc. State Collage, PA, USA.

Montaño-Mata, N. J. y J. C. Núñez (2003). Evaluación del efecto de la edad de transplante sobre el rendimiento en tres selecciones de ají dulce *Capsicum chinense* Jacq. en Jusepín. Estado Monagas. *Revista de la Facultad de Agronomía* 20, pp. 144-155.

Montgomery, D. C. (1991). *Design and analisys of experiments*. Tercera edición, Wiley, Nueva York.

Montgomery, D. C. y E. A. Peck (1991). *Introduction to linear regression Analysis*. Wiley, Nueva York.

Myers, R. H. y Montgomery, D. C. (1995). *Response surface methodology: Process and product optimization using designed experiments*. John Wiley, Nueva York.

Myers, R. H., Montgomery, D. C. y Vining, G. G. (2002). *Generalized linear models: With applications in engineering and the sciences*. John Wiley, Nueva York.

Nair, V. N. (1992). "Taguchi's parameter design: A panel discussion". *Technometrics*, 34, pp. 127-161.

Nair, V. N. y Pregibon, D. (1986). "A data analysis strategy for quality engineering experiments". *AT&T Technical Journal*, 65, pp. 73-84.

Phadke, M. S. (1989). *Quality engineering using robust design*. Prentice-Hall, Englewod Cliffs, N. J.

Phadke, M. S., Kakar, R. N., Speedey, D. V. y Greico, M. J. (1983). "Offline quality control in integrated circuit fabrication using experimental design". *The Bell System Technical Journal*, 62, pp. 1273-1310.

Pinal, L., Cedeño, M., Gutiérrez, H. y Alvarez-Jacobs, J. (1997). "Fermentation parameters influencing higher alchohol production in the tequila process". *Biotechnology Letters*, 19, pp. 45-47.

Potcner, K. J. y Kowalski, S. M. (2004). *How to analyze a split-plot experiment*. Quality Progress 37, núm. 12, pp. 67-74.

Ramírez, J., Gutiérrez, H. y Gschaedler, A. (2001). "Optimization of astaxanthin production by Phaffia rhodosyma through factorial design and response surface methodology". *Journal of Biotechnology*, 88, pp. 259-268.

Salcedo, Ch. B., Osuna, C. J., Guevara, L. F., Domínguez, J. y Paredes, L. O. (2002). Optimization of isoelectric precipitation method to obtain protein isolates from amaranth seeds. *J. Agric. Food Chem.* 50, 6515-6520.

SAS Institute Inc., SAS/QC Software. ADX Menu System for Design of Experiments, Version 6. 1a. ed., Cary, NC: SAS Institute Inc., 1994.

Statgraphics (Statistical Software). StatPoint, Inc., Herndon, Virginia, USA.

Steel, R. G. D. y Torrie, J. H. (1980). *Principles and Procedures of Statistics: A Biometrical Approach*. 2a. ed., McGraw Hill International Editions.

Taguchi, G. (1987). *System of experimental design*, vols. I y II. UNIPUB, Nueva York.

Technometrics **1**, 311-341.

Technometrics, **30**(1), pp. 1-17.

Vargas del Río, D., Zárate del Valle, P. F. y Gutiérrez-Pulido, H. (2005). La Actividad Puzolánica del las Tobas Pumíticas de la Zona Metropolitana de Guadalajara. *e-Gnosis* [online], vol. 3, art. 9, 15 páginas. (www.e-gnosis.udg.mx/).

Vargas Marin, G., Urzúa, D. y Gutiérrez Pulido, H. (2006). Influencia de la mezcla y del curado en la calidad del tabicón elaborado en la Zona Metropolitana de Guadalajara. *Ingeniería*, vol. 10, núm. 1, pp. 25-38.

Villafranca, R. R. y R. L. Zúñica (1993). *Estadística: diseño de experimentos, modelos de regresión*. Universidad Politécnica de Valencia, Valencia, España.

Weihs, C., Berres, M. y Grize, Y. (1995). "Statistical design of experiments in industrial practice". *Surveys on mathematics for industry*, 5, 49-73.

Índice analítico

A

Aberración mínima, 280
  diseños fraccionados con, 280
Aditivos, 304
Aleatorización, 13
  restricciones de, 108
Algoritmo de Yates, 231
Alias structure, 286
Análisis, 328
Análisis canónico, 399
  pasos del, 399
Análisis de cordillera, 406
  gráfica de, 408
  método de, 407
Análisis de regresión, 338-383
  calidad del ajuste en el, 351
  estimación y predicción por intervalo, 357
  intervalos de confianza, 371
  pruebas de hipótesis, 346-351, 365
  regresión lineal múltiple, 360
  regresión lineal simple, 340
Análisis de varianza, 56-96, 104
  experimentos con un factor, 56-96
    diseño completamente al azar, 62-73
    familia de diseños para tratamientos, 62-65
    pruebas de rango múltiple, 74-81
    tamaño de la muestra, 89-98
    verificación de supuestos del modelo, 81-89
Análisis del diseño factorial 2^3, 192
Ancho de los niveles, 332
ANOVA, 11, 62
ANOVA desglosado, 248
  pasos para llegar al, 171
Aprendizaje, proceso interactivo de, 6
Arreglo externo, 307
Arreglo factorial, 11
Arreglo interno, 307
Arreglo interno-externo en parcelas divididas, 469
Arreglos ortogonales, 303
Aumentar el experimento, 334
Autocorrelación, 354
Azar, 20

B

Balanceado, diseño, 398
Bloque, 102-105
  completos al azar, 102
  diseños en, 183
Bloqueo, 13
Bloqueo, principio de, 62
Bloques, 216
  casos de, 217
Bloques completos al azar, 102, 416
Bloques con contrastes, 218
Bloques con réplicas, 217
Bloques incompletos balanceados, 416
Búsqueda de primer orden, 389
Búsqueda de segundo orden, 389
Búsqueda I , 389
Búsqueda II, 389

C

Cálculos manuales, 71
Calidad, 4-6
  control de, 5
  diseño de experimentos en, 5
    en Japón, 6
    evolución, 5
  problema de, 10
Calidad del producto, 296
Candidato a óptimo, 400
CEP, *véase* Control estadístico de procesos
Ciclo de Deming, 322
  pasos del, 322
Ciclo de la calidad, 335
Cociente señal ruido, 307
Coeficiente de correlación, 332
Coeficiente de correlación múltiple, 368
Coeficiente de determinación, 178, 351
Coeficiente de determinación ajustado, 351

Coeficiente de regresión, 361
Coeficiente lineal β_i, 492
Colapsar, 332
Colapsar el diseño, 199
Columnas de signos para contrastes, 193
Columnas de un efecto de interacción, 193
Comparación de cuero, 63
Comparación de dos centrifugadoras, 41
Comparación de dos máquinas, 39
Comparación de dos tratamientos, 39
Comparación de medias, 138
Comparación por contrastes, 179
Comparaciones de rangos múltiples, 74-81
Completo, 102
Componentes de varianza, 154
Comprobación de supuestos, 113
Concepto de robustez, 296
Condiciones de proceso, 8
Conocimiento no estadístico, 12
Construcción de fracciones, 265
Contraste, 79, 171
    definición, 170
Contrastes ortogonales, 80
Control de calidad, 4-6
    diseño de experimentos en el, 5
    en Japón, 6
        evolución del, 5
Control de calidad fuera de línea, 5, 296
Control estadístico de calidad, 5
Control estadístico de procesos, 4
Control total de calidad, 6
Cordillera estacionaria, 405
Cordillera óptima, 406
Corridas experimentales, 323
Cresta ascendente, 405
Cresta descendente, 405
Cribado, 389
Cuadrado medio de tratamiento, 68
Cuadrado medio del error, 68, 346
Cuadrados del error, 68
Cuadrados medios, 68, 136
Cuadro, 109
Cuadro grecolatino, 115
    diseño en, 115
Cuadro latino, 109
    diseño en, 109
Cuadro latino estándar, 114
Curvas de nivel, 179, 388
Curvatura pura, 242, 334

D

Datos aberrantes, 182
DBCA, *véase* Diseño de bloques completos al azar
DCA, *véase* Diseño completamente al azar
Decisión equivocada, 33
Deseabilidad global, 441, 442
Desviación estándar, 24
Deterioro, 302
Diagrama de caja, 70, 72
Diagrama de Pareto, 197
Diagrama esquemático, 302
Diagramas de cajas simultáneas, 72
Diferencia mínima significativa, 74
Dirección del aprendizaje, 6
Dirección óptima, 421
Diseño, 10
Diseño anidado o jerárquico, 454
Diseño ANOVA, 65
Diseño balanceado, 64
Diseño completamente al azar, 65
Diseño de Box-Benhken, 239
Diseño de experimentos, 2-23
    clasificación, 14, 15
    control de calidad, 5
    definiciones, 6
    en Japón, 6
    en la industria, 4
    etapas, 10
    investigación, 5
    métodos estadísticos, 12
    principios básicos, 12
    problemas de la industria, 4
    selección, 14
Diseño de parámetro, 296
Diseño de primer orden, 415
Diseño de resolución III, 265
Diseño de resolución IV, 265
Diseño de resolución V, 265
Diseño de vértices extremos, 496
Diseño en bloques, 99-125
Diseño en cuadro grecolatino, 115
Diseño en cuadro latino, 109-114
    ANOVA para el, 110
    fuente de variabilidad, 109
Diseño en parcelas divididas, 460
Diseño estadístico de experimentos, 4
Diseño factorial 2^2, 128, 168
Diseño factorial 2^3, 183, 238
Diseño factorial 2^k, 168
    diseño factorial 2^2, 168

diseño factorial 2^3, 185
diseño factorial 2^k no replicado, 195
experimento 2^2, 173
experimento 2^5 no replicado, 260
experimento factorial 2^3, 192
factorial 2^k con punto al centro, 214
factorial 2^k en bloques, 216
Diseño factorial 2^{k-1}, 261, 265
Diseño factorial 2^k no replicado, 195
Diseño factorial fraccionado 2^{k-1}, 261
Diseño factorial general, 150
Diseño factorial general 2^k, 192
Diseño ortogonal, 413
Diseño robusto, 299-317
arreglos ortogonales, 303
cociente señal ruido, 309
concepto de robustez, 298
diseño con arreglo interno y externo, 307
diseño de parámetros, 307
experimento robusto, 310
factores de control, 299
filosofía Taguchi, 296
Diseño robusto de parámetros, 296
Diseño rotable, 413
Diseño simplex, 415
Diseño simplex con centroide, 486
Diseño simplex-reticular $\{q, m\}$, 486
Diseños de Plackett-Burman, 281
Diseños en bloques completos al azar, 102
Diseños experimentales, 14
Diseños factoriales, 126-164
conceptos básicos en, 128-132
de dos factores, 134-143
de tres factores, 143-148
diseño factorial general, 150
experimentación factorial *vs*. mover un factor a la vez, 132
modelo de efectos aleatorios, 153
transformaciones para estabilizar varianza, 148
ventajas de los, 133
Diseños factoriales de dos factores, 134-148
Diseños factoriales de tres factores, 143-148
Diseños factoriales fraccionados 2^{k-2}, 269
Diseños factoriales fraccionados 2^{k-p}, 257-272
concepto de resolución, 260
construcción de fracciones, 2^{k-1}, 265
diseños factoriales fraccionados 2^{k-2}, 269
diseños factoriales fraccionados^{k-p}, 273
experimentos 2^{5-1}, 267
experimentos 2^{7-4}, 276
Diseños para comparar tratamientos, 62-66
Distribución de probabilidad e inferencia estadística, 20-23
Distribución *F*, 22
Distribución normal, 21
Distribución *T* de Student, 21
Distribuciones ji-cuadrada, 21

E

Ecuación canónica, 402
Ecuación determinante, 405
Efecto *A*, 184
Efecto *ABC*, 219
Efecto *B*, 169
Efecto de bloque, 108, 219
Efecto de curvatura, 238
Efecto de interacción, 109, 129, 175
Efecto de un factor, 129
Efecto lineal y cuadrático, 241
Efecto térmico, 186
Efectos aleatorios, modelo de, 153
Efectos alias, 261
estructura, 264
interpretación de, 264
Efectos cuadráticos puros, 401
Efectos de interacción, 62
Efectos del error, 196
afectan claramente, 197
intermedios, 197
no afectan claramente, 197
Efectos fijos, 65
Efectos generadores, 270
Efectos principales, 129
Efectos significativos, 172
Eigenvalores, 405
Enfoque al cliente, 296
Enfoque al fabricante, 296
Entre más grande mejor, 306
Entre más pequeño mejor, 306
Equivalentes, 37
aceptación de, 38
rechazo de, 38
Error aleatorio, 9, 62
Error de parcela, 467
Error de subparcela, 467
Error estándar, 30, 346
Error estándar de estadístico, 188
Error estándar de estimación, 346, 352
Error experimental, 9

Error puro, 215
Error tipo I, 33
Error tipo II, 33
Escalamiento ascendente, 394
Escalamiento descendente, 394
Estadística, 20
  inferencia, 20
    elementos de la, 18-59
Estadístico, 20, 21
  parámetro y, 20
Estadístico apropiado, 154
Estadístico de Durbin Watson, 355
Estadístico de prueba, 31, 37
  propiedades del, 31
Estadísticos de pruebas apropiadas, 154
Estadísticos señal ruido, 309
Estimación por intervalo, 23
Estimación puntual, 23
Estimación puntual y por intervalo, 23
Estimación, 20
Estimación, error estándar de, 346
Estructura de diseño, 464
Estructura del alias, 263
Estudio R&R, 155
Evolución del control de calidad, 5-7
  diseño de experimentos en el, 5
Experimentación, 320
Experimentación a prueba y error, 12
Experimentación factorial *vs*. mover un factor, 132
Experimentación factorial, 132
Experimentación secuencial, 12
Experimentar, 5
Experimento, 7, 220, 276
Experimento 2^2, 173
Experimento de Ossie Cadenza, 322
Experimento factorial, 128
Experimento mixto, 500
Experimento ortogonal, 413
Experimento robusto, 310
Experimentos con mezclas, 482
Experimentos con un solo factor, 60-98
  comparaciones de rangos múltiples, 74-81
  diseño completamente al azar, 62-73
  diseño de ANOVA, 62-73
  elección del tamaño de muestra, 89-98
  familias de diseños, 62-65
  verificación de supuestos del modelo, 81-89
Experimentos con uno y dos tratamientos, 18-59
  distribución de probabilidad, 21
  estimación puntual y por intervalo, 23-29
  hipótesis para dos medias, 39
  igualdad de varianzas, 43
  poblaciones pareadas, 44-48
  prueba de hipótesis, 29
  prueba para la media, 34
  prueba para la varianza, 36-39
Experimentos, diseño de, 2-16
Exponentes *s* y *t*, 442

F

Factor *A*, 184
Factor bloqueado, 13
Factor cualitativo, 128
Factor cuantitativo, 128
Factor de ajuste, 301
Factor de flux, 320
Factor de interés, 62
Factor dominante, estimación de, 283
Factor para residuos, 87
Factor señal, 301
Factores, 8, 324
Factores aleatorios, 153
Factores controlables, 8
Factores cruzados, 455
Factores de bloque, 62
Factores de bloque y ruido, 330
Factores de estudio, 10
Factores de ruido, 8, 299
  controlables, 299
  no controlables, 299
Factores estudiados, 8
Factores fijos, 153
Factorial aleatorio, 457
Factorial anidado, 455
Factorial fraccionado, 279
Factorial fraccionado saturado, 281
Factorial mixto, 247
Factoriales 2^k con punto al centro, 214
Factoriales 2^k en bloques, 216
Falta de ajuste, 215
Familia de diseños factoriales 2^k, 192
Familia de diseños para comparar tratamientos, 63
Filosofía Taguchi, 296
  metas, 296
  propiedades, 296

Forma canónica, 399
Fórmula para intervalo de confianza, 29
Fórmula para prueba de hipótesis, 46
Fracción a la mitad, 261
Fracción complementaria, 262
Fracción principal, 262
Fuentes de variabilidad, 62, 102
Fuera de línea de producción, 5
Función de costo social, 296
Función de deseabilidad, 441
Función de pérdida, 296

G
Generador del diseño, 262
Grabado mesa, 185
Grados de libertad, 22
Gráfica de contornos, 472
Gráfica de diagnóstico, 92
Gráfica de efectos en papel normal, 210
Gráfica de factor contra residuos, 86
Gráfica de medias, 70, 73
Gráfica de probabilidad en papel normal, 83
  pasos para la, 83
Gráfica de probabilidad normal, 83
Gráfica de residuos, 86, 92
Gráfico de Daniel, 197
Gráfico de Pareto, 202
Gráfico de superficie, 179
Gráfico de trazas, 493
Grecolatino, cuadro, 115

H
Half normal, 197
Hiperboloide, 392
Hiperelipsoide, 392
Hiperplano, 392
Hipersuperficie, 437
Hipótesis, 29-34
  alternativa, 30
    bilateral, 30
    unilateral, 30
  fórmulas para procedimiento de, 49
  nula, 30
  procedimientos de prueba, 49
  prueba de, 20, 29
    concepto, 29
Hipótesis alternativa, 30
Hipótesis alternativa bilateral, 30
Hipótesis alternativa unilateral, 30
Hipótesis de autocorrelación negativa, 354
Hipótesis de interés, 145
Hipótesis estadística, 30
Hipótesis nula, 30
Hipótesis para dos medias, 39

I
Independencia, 88
Independencia de residuos, 329
Industria, 4
  diseño de experimentos en la, 4
Inferencia estadística, 18-59
  elementos de la, 18
    distribución de probabilidad, 21-23
    equivalentes, 37
    estimulación por intervalo, 23-26
    estimulación puntual, 23
    fórmula de prueba de hipótesis, 49
    hipótesis para dos medias, 44
    igualdad de varianzas, 43
    parámetros y estadísticos, 20
    poblaciones pareadas, 44
    población y muestra, 20
    prueba de hipótesis, 29-34
    prueba para la media, 34
  estimación, 20
  prueba de hipótesis, 20
Interacción, 270
Interacción entre dos factores, 65
Interacción triple, 183
Interpretación, 11, 328
Interpretación de efectos activos, 146, 220
Intervalo de aceptación, 31
Intervalo de confianza, 24, 39, 91
  fórmulas para, 28
  longitud de, 25
  resumen de, 29
Intervalo de confianza para una media, 25
Intervalo de confianza y predicción, 371
Intervalo de rechazo, 31
Intervalo para la proporción, 28
Intervalo para la recta de regresión, 357
Intervalo para la varianza, 27
Introducción al diseño robusto, 294-317
  arreglos ortogonales, 303
  cociente señal ruido, 309
  concepto de robustez, 298
  diseño con arreglos, 307
  diseño de parámetros, 307
  experimento robusto, 311
  factores de control, 312
  filosofía Taguchi, 296

Investigación, 5
diseño de experimentos y, 5-6
Isolíneas, 179, 338

J
Japón, 6
calidad en, 6
control de calidad en, 6
diseño de experimentos en, 6
evolución del control de calidad, 6-7

K
k-ésimo tratamiento, 79

L
Latino, 109
diseño en cuadro, 109
Línea de producción, 6
Longitud del intervalo de confianza, 25

M
Magnitud del efecto, 198
Matriz de diseño, 92
Matriz de diseño aleatorio, 200
Media global, 65
Media(s), 34-36
hipótesis para dos, 39
comparación de dos tratamientos, 39
prueba para la, 4
Medición del error absoluto, 352
Mediciones individuales, 179
Mejor corte, 439
Mejor tratamiento, 387
Método de Duncan, 77
Método de Dunnet, 78
Método de mínimos cuadrados, 342
Método de Sheffé, 80, 81
Método de Tukey, 76
Método del valor crítico, 350
Método gráfico, 437
Método gráfico de optimización simultáneo, 420
Método LSD, 74
Método Taguchi, 296
función de pérdidas, 297
metas, 297
objetivo del, 296
Metodología de caracterización de proceso, 323
Metodología de superficie de respuesta, 333
elementos de la, 391
Métodos de comparación múltiples, 70
Métodos de ensamble, 63
Métodos estadísticos, 12
Mínima aberración, 280
Mínimos cuadrados, método de, 342
Modelo, 389
Modelo de ANOVA, 82
supuestos del, 81
independencia, 81
normalidad, 81
varianza constante, 81
Modelo de efectos, 153
Modelo de efectos aleatorios, 153
Modelo de efectos fijos, 65
Modelo de primer orden, 394
Modelo de regresión, 348
ANOVA del, 334
Modelo de superficie de respuesta, 412
Modelo exponencial, 340
Modelo jerárquico, 435
Modelo logarítmico, 340
Modelo mixto, 156
Modelo multiplicativo, 340
Modelo recíproco, 340
Mover un factor a la vez, 132
Muestra, 20, 89-98
elección del tamaño de la, 89
población y, 20
Muestra por intervalo de confianza, 91
Muestra representativa, 20
Muestras pareadas, 44
Muestreo aleatorio, 21
Múltiples, métodos de comparación, 70
Multiplicación módulo, 189

N
Nivel de un factor, 325
Niveles, 7
Niveles del factor, 63
Normalidad, 83
Notación 2^{k-1}, 261
Notación de puntos, 66
Notación de Yates, 169, 171
Nulificar, 13

O
Objetivo del experimento, 14
clasificación del, 14
Obleas, 185
análisis de Pareto en, 185
Obleas de silicio, 185

Operación evolutiva, 389
Operador, 13
Oportunidades de mejora, 5
Optimización, 389
Optimización de escalamiento, 394
  pasos de la, 394
Optimización de neumáticos, 407
Optimización de procesos con superficie de respuesta, 384-430
  concepto de optimización, 389
  diseño de superficie de respuesta, 409
  metodología, 389
  modelos, 393
  optimización simultánea, 420
  técnica de optimización, 393
Optimización en dos pasos, 309
Optimización simultánea, 420
Óptimo individual, 434
Óptimo simultáneo, 434
Orden completamente al azar, 45
Orden de corrida, 92
Ortogonalidad, 303, 419
  propiedad de, 303

P

Papel de probabilidad normal, 210
  gráfica de efectos en, 210
Parámetro de regresión lineal, 348
Parámetros, 296
  estadísticos y, 20
Parámetros del proceso, 8
Parcela, 460
Parcelas doblemente divididas, 464
Parcelas en bloques, 467
Parejas de medias de tratamientos, 74
  comparación de, 74
Pareto estandarizado, 187
Parte lineal del efecto *A*, 243
Patrón de longitud de palabra, 280
Patrón tipo corneta, 205
Peso de costales, 35
Pesos w_i (importancia relativa), 442
Planeación, 10, 324
  actividades, 323
  diseño, 324
  error, 323
  etapas, 323
Planeación de un experimento, 318-336
  actividades, 323
  ciclo de Deming, 322
  control de factores, 330
  etapas, 323
  experimentación, 320
Planeación del experimento, 186
Planear un experimento, 323
Planteamiento de hipótesis estadística, 30
Población, 20
  finita, 20
  infinita, 20
  muestra y, 20
Población finita, 20
Población infinita, 20
Poblaciones pareadas, 44-48
Polinomio, 405
Porción axial, 417
Porción factorial, 417
Potencia, 74
Potencia de la prueba, 33
Precisión uniforme, 418
Predicción, 179
Predicción del mejor tratamiento, 178
Predichos, 190
Prevención, 5
Principio de aleatorización, 330
Principio de bloqueo, 62, 174
Principio de Pareto, 152
Probabilidad, distribuciones de, 21
Probabilístico medio normal, 197, 210
Problemas de calidad, 9
Problemas de mezclas, 483
Procedimientos de prueba de hipótesis, 49
Proceso de deducción, 5
Proceso de inducción, 5
Proceso interactivo de aprendizaje, 5
Proceso interactivo de la experimentación, 6
Propiedad de ortogonalidad, 262
Proyectar el diseño, 199
Prueba de Bartlett de homogeneidad de varianzas, 87
Prueba de Durbin Watson, 354
Prueba de falta de ajuste, 355-401
Prueba de hipótesis, 20, 29-34
  concepto, 29
  criterio de rechazo, 31
  errores tipo I y II, 33
  estadística de prueba, 31
  fórmulas para procedimientos de, 49
  planteamiento de hipótesis, 30
Prueba de rangos múltiples, 74
Prueba de Shapiro-Wilks para normalidad, 85
  pasos para la, 85
Prueba *F* del ANOVA, 69

Prueba para la media, 34
varianza desconocida, 34
Prueba para la varianza, 36
Prueba para la variedad de varianzas, 43
Prueba, potencia de la, 33
Prueba *T* de Student, 69
Prueba y error, 6
Pruebas analíticas, 83
Pruebas bilaterales, 32
Pruebas de dos colas, *véase* Prueba bilateral
Pruebas de rangos múltiples, 74-81
Pruebas de una cola, *véase* Prueba unilateral
Pruebas gráficas, 83
Pruebas unilaterales, 32
Punto central, 214
Punto estacionario, 400
Punto óptimo, 387
máximo, 392
mínimo, 392
Punto óptimo gráfico, 441
Punto silla, 404
Puntos axiales, 330
Puntos de diseño, 8
p-value, 69

R

Rangos múltiples, 74
pruebas de, 74
Razón señal ruido, 307
Reactivo, 238
Recta de regresión, 357
Reescalar, 332
Región, 31
Región de aceptación, 31
Región de operabilidad, 386
Región de rechazo, 31
Región experimental, 169, 386
Región factible, 438
Regresión lineal múltiple, 360
análisis de, 389
modelo de, 360
pruebas de hipótesis, 365
selección de variables, 368
Regresión lineal simple, 340
Regresión óptima, 357
Regresión simple, 357
estimación y predicción, 357
Relación definidora, 263, 270, 273
Relación modelo diseño, 392
Rendimiento eléctrico, 185
Rendimiento mecánico, 185
Rendimiento óptimo, 388
Repetición, 13
Repeticiones, 173
Repeticiones al centro, 391
Réplica, 134
Replicar, 332
Representación del efecto de interacción, 130
Representación geométrica, 169
Representación geométrica del diseño, 262
Residual, 81, 180
Residuos, 81, 344
desviación estándar de, 181
independientes, 182
normales, 182
orden de corrida, 182
varianza constante, 182
Residuos asociados a la observación, 82
Residuos en papel normal, 190
Resolución, 265
alta, 265
definición, 265
diseños, 265
comentarios sobre la, 279
Resolución III, 276
Restricción de aleatorización, 108
Robustez, 296
idea de, 297
tipos de estudio de, 306
Robusto, 298
Ruido externo, 302
Ruido interno, 302

S

Saturado, 332
diseño altamente, 331
Seudocomponentes, 499
Significancia calculada, 38
Significancia de efectos, 208
menos clara, 208
Significancia estadística, 12
Significancia observada, 38
Significancia predefinida, 38
Simplex, 483
Soldadura de ola, 320
SPC, *véase* Control estadístico de procesos
Subparcela, 460
Suma de cuadrados de regresión, 348
Suma de cuadrados de tratamientos, 67
Suma de cuadrados del error, 67, 193, 345
pasos para construir la, 196

Suma de observaciones en cada tratamiento, 168
Suma total de cuadrados, 67
Superficie de respuesta, 179, 333, 384, 393
  diseños de, 396
  metodología, 384
  modelos, 396
  propiedades, 414
Superposición de superficies, 437
Suposición de varianzas desconocidas, 40
Supuesto de independencia, 354
  verificación de, 354

T
T de Student, 189
Tabla de análisis de varianza, 69
Tabla de ANOVA, 69
Tabla de distribución *F*, 513
Tabla de distribución normal, 510
Tabla de prueba de independencia, 520
Tabla de rango estudentizado, 515
Tabla de signos, 171, 174
Tabla para comparación de medias Duncan, 555
Tabla *T* de Student, 512
Tablas de Dunnet, 518
Taguchi, 299-317
  arreglos ortogonales, 303
  cociente señal ruido, 307
  diseño con arreglo, 306
  diseño de parámetros, 307
  experimento robusto, 310
  factores de control, 299
  filosofía, 296
  robustez, 298
Tamaño de broca, 173
Técnica *backward*, 371
Técnica *forward*, 371
Temperatura de grabado, 190
Tiempo de ensamble, 63
Total de observaciones en cada tratamiento, 168
Trabajo experimental, 328
  planear el, 328
Transformaciones para estabilizar varianza, 148-150
Traslación, 403
Tratamiento con punto central, 214
Tratamiento control, 78
Tratamiento de referencia, 63
Tratamiento elegido, 333
Tratamientos, 8, 39
  comparación de dos, 39
  hipótesis para dos medias, 39

U
Unidad experimental, 7, 461
Unidades codificadas, 174
Unidades originales, 173

V
Valor crítico, 37
Valor de *p-value*, 172
Valor nominal mejor, 309
Valores característicos, 403
Valores óptimos, 386
Valores propios, 403
Variable aleatoria, 21
Variable de respuesta, 8, 10, 340
Variable independiente, 340
Variable regresora, 340
Variables, 7
Variables canónicas, 399
Variables de diseño, 8
Variables de entrada, 8
Variables de salida, 326
  actividades de, 326
Varianza, 36
  igualdad de, 43
  prueba para la, 36
  transformaciones para estabilizar, 148
Varianza constante, 85, 329
Varianzas desconocidas, 40
  suposición de, 40
Verificación de los supuestos del modelo, 81-89
Vibración de la ranuradora, 173

Y
Yates, 168